# REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-99-

0057



| | | |
|---|---|---|
| 1. AGENCY USE ONLY *(Leave Blank)* | 2. REPORT DATE<br>1 Feb 99 | 3. REPORT TYPE A... COVERED<br>Final 1 Dec 97 to 30 Nov 98 |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>Psychomotor and Perceptual Speed Abilities and Skilled Performance | 5. FUNDING NUMBERS<br>F49620-98-1-0036<br>2313/BS<br>61102F |
| 6. AUTHORS<br>Phillip L. Ackerman | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Georgia Tech Research Corporation<br>400 Tenth Street N.W.<br>Atlanta, Georgia 30332-0420 | 8. PERFORMING ORGANIZATION REPORT NUMBER<br>Ackerman-01 |
| 9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>Air Force Office of Scientific Research<br>AFOSR/NL<br>801 North Randolph Street, Room 732<br>Arlington, VA 22203-1977 | 10. SPONSORING / MONITORING AGENCY REPORT NUMBER<br>F49620-98-1-0036 |

11. SUPPLEMENTARY NOTES
This report supersedes report of 12/15/97 from the University of Minnesota covering project from 2/14/96 - 12/15/97.

| | |
|---|---|
| 12a. DISTRIBUTION / AVAILABILITY STATEMENT<br>Unlimited | 12b. DISTRIBUTION CODE |

13. ABSTRACT *(Maximum 200 words)*

The research in this project was aimed at three broad approaches to development and assessment of psychomotor and perceptual speed ability predictors of skilled performance. The first approach takes advantage of computerized touch-panel devices for assessment of a series of psychomotor abilities. The second approach links individual differences in psychomotor abilities with perceptual speed abilities -- which have been shown to be important predictors of the acquisition of skilled performance. The third approach evaluates the new test battery for predicting individual differences in task performance. For this project, seven touch-panel computerized psychomotor tests were developed, as follows: (1) Tapping and Alternate Tapping; (2) Choice Reaction Time; (3) Serial Reaction Time; (4) Maze Tracing; (5) Mirror Tracing; (6) Maze Pursuit; and (7) Rotary Pursuit. These new psychomotor tests, along with a taxonomically derived set of perceptual speed tests, were subjected to empirical evaluation in a series of experiments. The new tests show substantial promise in accounting for important sources of performance variance, as indicated from validation with basic skill and complex task performance criteria.

| | |
|---|---|
| 14. SUBJECT TERMS<br>Psychomotor<br>Individual Differences<br>Abilities | 15. NUMBER OF PAGES<br>80<br>16. PRICE CODE |

| | | | |
|---|---|---|---|
| 17. SECURITY CLASSIFICATION OF REPORT<br>Unclassified | 18. SECURITY CLASSIFICATION OF THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION OF ABSTRACT<br>Unclassified | 20. LIMITATION OF ABSTRACT<br>UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-1
298-102

Psychomotor and Perceptual Abilities and Skilled Performance

FINAL REPORT
(Period 12/1/97 - 11/30/98)

Phillip L. Ackerman

School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170

Submitted to:

John F. Tangney
AFOSR/NL
Arlington, VA 22203-1977



This is a Final Report prepared for the Air Force Office of Scientific Research.

# SUMMARY

The research program is reviewed in the context of the literature and previous AFOSR/AFHRL contract support. The main thrust of the earlier Air Force contracts (1985-1987) was the development and validation of a theory of the cognitive ability determinants of individual differences in skill acquisition (e.g., Ackerman, 1988, 1990). The focus of the more recent AFOSR/AFHRL contracts (1988-1991 and 1993-1996) has been to broaden the basic theory and empirical foundation to encompass complex cognitive tasks (including planning, problem solving, and decision making) and an integration of cognitive and self-regulatory processes as they determine individual differences in skill acquisition (e.g., see Ackerman, 1992; Ackerman & Kanfer, 1993, Ackerman, Kanfer & Goff, 1995; Ackerman & Kyllonen, 1991; Ackerman & Woltz, 1994; Kanfer & Ackerman, 1989, 1990, 19961.

The research in the current project focuses on three broad approaches to development and assessment of psychomotor and perceptual speed ability predictors of skilled performance: The first approach takes advantage of computerized touch-panel devices for assessment of a series of psychomotor abilities; The second approach links individual differences in psychomotor abilities with perceptual speed abilities, which have been shown to be important predictors of the acquisition of skilled performance. The third approach evaluates the new test batteries for predicting individual differences in task performance. For this project, seven touch-panel computerized psychomotor tests were developed: (1) Tapping and Alternate Tapping; (2) Choice Reaction Time; (3) Serial Reaction Time; (4) Maze Tracing, (5) Mirror Tracing, (6) Maze Pursuit, and (7) Rotary Pursuit. These new psychomotor tests along with a taxonomically derived set of perceptual speed tests were subjected to empirical assessment in a series of experiments. The new tests show substantial promise in accounting for important sources of performance variance, as indicated from validation with basic skill and complex task performance criteria.

Table of Contents

## Psychomotor and Perceptual Abilities and Skilled Performance

### I. Overview

This research project has three components, briefly reviewed below. The first component of the project involved the use of general purpose desktop computers and touch-panel displays to generate a software suite that assesses individual differences in psychomotor skills and aptitudes. The suite includes seven test types that assess: (1) Tapping and Alternate Tapping; (2) Choice Reaction Time; (3) Serial Reaction Time; (4) Maze Tracing, (5) Mirror Tracing, (6) Maze Pursuit, and (7) Rotary Pursuit. The tests have the advantage of ameliorating the four main obstacles that have historically prevented wide-spread use of psychomotor testing, namely: (1) Fabrication costs (the current project involved off-the shelf equipment); (b) Calibration requirements (the new technology required minimal time and effort for calibration); (c) Examiner training (minimal training of examiners was required, given the simplicity of operation and limited needs for adjustment and maintenance); and (d) Low examiner-to-examinee ratios (because the systems are highly self-contained, and include on-board intelligence for upkeep). Within this portion of the project, the test battery was developed and subjected to empirical assessment of reliability and validity (e.g., correlations with other psychomotor ability assessments).

Previous AFOSR projects have involved an extensive investigation of perceptual speed abilities, including the development of a taxonomic representation of these domains. In several studies it has become clear that tests of complex perceptual speed abilities are important predictors of individual differences in learning and skilled performance. From a theoretical perspective, these investigations are important because they shed light on a domain of human abilities that is not well-understood. From a practical perspective, this work has demonstrated that substantial gains in the prediction of training success and performance can be accomplished by the proper selection of appropriate perceptual speed measures. For example, Ackerman and his colleagues demonstrated that two measures in particular (a variation of the old Army Air Force Dial Reading Test, and an FAA-inspired Directional Headings Test) provide substantial incremental validity in predicting performance of air traffic controllers, both in the laboratory and in the field. Moreover, empirical work (Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995) has shown that complex perceptual speed tests also capture aspects of personality and self-regulatory processes that may interfere with training success on complex skill tasks. Research was also conducted that suggests perceptual speed abilities form a linkage between higher-level cognitive and intellectual abilities on the one hand, and psychomotor abilities on the other hand. Thus, when attempting to integrate new measures of psychomotor abilities into the broader aptitude context, it will be necessary to delineate how psychomotor and perceptual speed abilities relate to one another, and determine their respective contributions to the prediction of learning and skilled performance. The second component of the project focused on three approaches to integrating perceptual speed and psychomotor abilities, namely: (a) Refinement of the taxonomy of perceptual speed abilities; (b) Development of a faceted battery of paper and pencil perceptual speed tests; and (c) joint examination of psychomotor

and perceptual speed ability factors.

The first two components of the research program provided descriptive statistics, reliability indices, and intercorrelations for a broad integrated battery of psychomotor and perceptual speed ability tests. In the third component of the research program, the integrated psychomotor/perceptual speed test batteries was evaluated against both basic skill acquisition tasks (i.e., the Kanfer-Ackerman Air Traffic Control Task©) and prediction of complex problem solving and decision making task performance (i.e., predicting performance on a high-fidelity air traffic controller simulation task called the Terminal Radar Approach Control [TRACON] simulation — see Ackerman, 1992; Ackerman & Kanfer, 1993; Ackerman, Kanfer & Goff, 1995).

Finally, this research program explored the potential for military applications, in that the neither the perceptual speed or psychomotor ability domains are well represented in the current Armed Services Vocational Aptitude Battery (ASVAB). Perceptual speed is only represented partly by the Numerical Operations test, and partly by the Clerical Speed test -- neither of which are sufficiently varied or complex to provide an incremental prediction of performance beyond that contributed by general ability. Psychomotor abilities are not represented in the ASVAB at all. Thus, incremental predictive validities from the new batteries of tests may be particularly useful in future revisions of selection procedures for the Air Force.

## II. Background

During the Second World War, when the U.S. Air Force was still the "Army Air Forces," (AAF) a large team of eminent psychologists led perhaps the most in-depth and productive series of research studies converging on nearly all aspects of human performance (ability testing, training, equipment design, force retention, and so on). Among these psychologists were R. L. Thorndike, J. Flanagan, J. P. Guilford, L. G. Humphreys, P. Fitts, and J. J. Gibson. Hundreds of ability tests, both paper and pencil and apparatus-based (including, for example, tests of dynamic spatial reasoning using motion pictures) were developed and subjected to empirical evaluation and refinement, with thousands of examinees. Selection batteries were created for pilots, navigators, bombardiers, gunners, and others. The 19 volumes of reports (*Army Air Force Aviation Psychology Program Research Reports, Vol 1-19*, Government Printing Office, 1947) describing this research enterprise still constitutes a major source of basic and applied research on cognitive/intellectual ability, perceptual speed ability, and psychomotor ability that clearly demonstrates the value of studying individual differences, as they affect the selection, training, and job/equipment design for humans in the context of complex task performance.

### Cognitive Abilities

In the decades subsequent to WWII, basic and applied research has progressed in myriad ways. Developments in information theory leading to information processing

psychology, and later cognitive psychology brought about new metaphors and paradigms for the determinants of human performance. Developments in intellectual ability theory (some of it predicated on the AAF work by Guilford and his colleagues (Guilford & Lacey, 1947; *AAF Report #5*) have led to substantially greater precision in specification of the nature of human cognitive and intellectual abilities. From an applied perspective, one source of progress in this area has focused on improved measures of general intellectual ability. Recent investigations (e.g., Earles & Ree, 1992; Ree, Earles, & Teachout, 1992) have shown that such measures have high validity for predicting performance across a wide array of military specialties. In addition, several streams of research have converged on additional ability - performance linkages, especially in the context of learning and skill acquisition (e.g., the four-source approach developed at the Air Force Armstrong Laboratory, see Kyllonen & Christal, 1989; Kyllonen, Tirre, Christal, 1991; Woltz, 1988). Thus, from a cognitive ability perspective, both general and content abilities, along with sources of information processing abilities have been shown to well-predict skilled performance.

## Perceptual Speed Abilities

Historically, the domain of perceptual speed abilities has not received the level of attention from researchers that general and broad content abilities have had. However, the importance of Perceptual Speed abilities for *complex cognitive task* performance emerged during WWII when the Army Air Force Aviation Psychology program found numerous perceptual speed measures that well-predicted pilot and navigator performance. Recent theoretical and empirical investigations have shown that perceptual speed abilities play a fundamental role in the development of skilled performance at nearly all levels of skills (from complex cognitive skills to fine motor skills). For example, in several studies with a high-fidelity air traffic controller simulation task (Terminal Radar Approach Control -- TRACON), a wide battery of ability tests (including general ability, spatial ability, and perceptual speed ability) was used to predict performance at various levels of skill (Ackerman, 1992; Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995). Two of the most valid tests for predicting performance throughout practice (but particularly at the more-skilled levels) were predominantly perceptual speed tests (namely, the Dial Reading Test and the Directional Headings Test, see Guilford & Lacey, *AAF Report #5*, 1947). These results are also consistent with data collected by investigators at the FAA Civil Aeromedical Institute in their prediction studies of real-world air traffic controllers (who typically undergo nearly three years of classroom and on-the-job training before they reach full proficiency levels).

Prior AFOSR-sponsored research has focused on the initial delineation and validation of a taxonomy of perceptual speed abilities (Ackerman & Rolfhus, 1996). This taxonomy will provide the foundation to build a battery of faceted tests for evaluation of their predictive validity for complex task performance. This work continued in the proposed research program, and was integrated with development of a battery of psychomotor ability tests.

Psychomotor Abilities

Tests of psychomotor skills and aptitudes have been an important part of the assessment spectrum developed for predicting individual differences in skilled performance over the past 80 years. While early tests were tailored for special applications, such as the prediction of performance for assembly-line workers and clerical operators (e.g., see Adkins et al., 1947; Hull, 1928; Münsterberg, 1913), more recent developments have shown that general-purpose psychomotor tests can be effective for predicting a wide variety of task performance criteria (for a mid-century review see Fleishman, 1953; also see Anastasi, 1982). However, while there have been attempts to make paper and pencil psychomotor tests, the major facet of successful psychomotor aptitude measures is that they are typically apparatus tests that is the tests typically require equipment (e.g., timers, electrical relays, lights, etc.) and/or tools of some type (e.g., pegboards, washers, levers, styluses, etc.).

Early factor-analytic work focused on refining a wide array of psychomotor tests to a small number of hypothetical common factors, such as in the domain of fine motor skills (e.g., see Seashore, 1940; Seashore, Buxton, & McCollom, 1940). However, the largest effort towards development and validation of psychomotor tests occurred during World War II, by the AAF. Under the direction of A. Melton (Melton, 1947), dozens of apparatus tests for psychomotor abilities were developed and evaluated. Tests were administered to more than 600,000 men in this program, and the tests were used to great success in selection and classification of aircrew personnel, starting in 1942. Psychomotor aptitude measures were clearly demonstrated to be valid predictors of complex-task performance (e.g., see Salvendy & Seymour, 1973).

However, Melton and his colleagues outlined four major obstacles in the use of these psychomotor tests:

(1) Fabrication costs. Each of the apparatus-based psychomotor tests had to be individually designed and fabricated, often to a high degree of precision, and at a high cost. There was no technological provision for a general-purpose psychomotor testing platform.

(2) Calibration requirements. Psychomotor tests using specialized apparatus require constant adjustment and calibration, to insure that the variance in examinee responses is attributable to individual differences in psychomotor ability, and not to differences across apparatuses.

(3) Examiner training. Use of specialized equipment requires a substantial amount of training of examiners.

(4) Examiner-to-examinee ratio. Melton and his colleagues found that it was typically necessary to have 1 examiner for every 4 examinees during psychomotor testing, in order to maintain proper supervision of the examinees and calibration/maintenance of apparatus.

In years subsequent to WWII, psychomotor apparatus tests were used as part of the general U.S. military selection battery, but substantial efforts were devoted to finding a method for eliminating the exorbitant equipment and personnel costs for administering psychomotor tests. One method was to construct paper-and-pencil psychomotor tests and compare the results with apparatus tests. Unfortunately, there turned out to be little overlap between such measures (e.g., see Fleishman, 1954, 1958; Melton, 1947), and this effort was largely abandoned. That is, paper-and-pencil tests of psychomotor abilities simply did not adequately predict individual differences in performance on tasks with substantial psychomotor demands.

Finally, in 1955, given the large costs associated with psychomotor apparatus tests, and the increasingly geographically distributed testing needs, the U.S.A.F. finally dropped these tests from the first-hurdle selection battery of pilots, bombardiers, and navigators, in favor an exclusively paper-and-pencil battery (see, e.g., discussion by Fleishman, 1956). Currently, such devices <u>are</u> mainly used for pilot selection in the Air Force [e.g., the Basic Attributes Test (BAT), see Carretta, 1987], as a much smaller number of examinees is involved, testing is centralized, and the consequences of selection errors are much more costly than entry-level selection.

Over the past two decades, psychological research (especially into the information-processing aspects of cognitive performance), has re-affirmed the critical importance of psychomotor abilities, especially in the prediction of individual differences in highly-skilled levels of performance (see Ackerman, 1987 for a review, also see Ghiselli, 1966). Ackerman's (1988) theory of the cognitive determinants of individual differences, and subsequent empirical investigations (e.g., Ackerman, 1990), emphasize that the final phase of skill acquisition is often well-predicted by individual differences in psychomotor abilities.

Even though it is clear from the historical record and current basic research findings that psychomotor abilities are important in selection and classification, until recently, there has been no technological solution to the logistic obstacles outlined by Melton and his colleagues during WWII. Two technological developments have provided the basis for a new approach to assessment of psychomotor abilities: First is the relatively inexpensive and pervasive general-purpose personal computers (PCs) with adequate graphic capabilities (e.g., 1024 x 768 pixels); Second is similarly inexpensive general-purpose touch-sensitive graphical display monitors (touch-panels). By linking these two technologies and designing a series of innovative software programs for the administration of psychomotor-skills stimuli and collection of responses, it may be possible to overcome the logistic obstacles to measurement of psychomotor abilities, for the enhancement of selection and classification applications for the U.S. Air Force.

Using PCs and touch-panel displays, all four of the obstacles to the use of psychomotor apparatus tests may be removed. That is, *fabrication* of special-purpose equipment is not necessary, as these are off-the-shelf items; *calibration* was expected to either be unnecessary, or take a matter of a few seconds per machine; *examiner training*

could be minimized, given the minimal needs for adjustment of the equipment, and the on-board diagnostic intelligence that can be built into the software; and *examiner-examinee ratios* may not be nearly so limited, given the ease of examinee-computer interaction, and the simple interface and relative familiarity of these touch-panel devices (e.g., the same technology is used in a variety of different public places, such as store kiosks, automated teller machines, state driver qualification stations, etc.). With additional refinements, these tests may very well be capable of examinee self-administration.

## Experiment Overview

The first two experiments in this research program were designed to evaluate the feasibility of psychomotor testing by computerized touchpanel monitors, in terms of traditional tactics of reliability and validity assessment. In the first study, the creation of three classes of psychomotor tests is described, namely: Tapping, Choice Reaction Time, and Serial Reaction Time. These tests were evaluated in terms of test-retest reliability and alternate-form reliability (finger vs. TouchPen® stylus input), validations against measures of simple and complex perceptual speed ability, and performance on a learning-task criterion (the Kanfer-Ackerman Air Traffic Control© simulation task). In the second study, two additional kinds of psychomotor tests were added to the battery (Maze Tracing and Mirror Tracing), and a greatly expanded battery of other abilities was administered, including measures of Spatial, Verbal, Numerical ability, and Mechanical Knowledge. The third study examined the final two touch-panel psychomotor tests (Maze Pursuit and Rotary Pursuit), along with apparatus psychomotor tests and a perceptual and psychomotor work-sample task. The fourth and final study examined the full suite of touch-panel psychomotor tests, along with cognitive and perceptual speed measures, in the context of predicting perceptual and motor task performance and performance on a decision making and problem solving task. In each of the first three experiments, the development of the new psychomotor tests is described and then followed by a review of the empirical results that evaluate the psychometric characteristics of these tests.

The first step for design of the psychomotor tests was to examine the extant test specifications from the literature (or from the actual apparatus tests), and to match stimuli as closely as possible to the apparatus tests. At that point, pilot testing was done to examine the feasibility of such designs — and then refinements were made to optimize the presentation and collection of responses. When the maze and mirror tracing tasks were designed, the starting point was with the original apparatus designs (the Lafayette maze tracing apparatus, and the Snoddy (1920) mirror star apparatus). However, in contrast to the apparatus tests, it was possible to generate multiple variations of each test on the computer — using similar number of turns and maze lengths, but different shapes.

## III. Experiment 1

Method

Participants. One hundred seventeen adults participated in this experiment. The examinees were recruited from around the campus of the University of Minnesota, with the

following criteria: (1) age between 18 and 30 years, (2) normal or corrected-to-normal vision, hearing, and motor coordination. Examinees were paid $60 for participating in Sessions 1 - 3, and if they participated in Session 4 (the apparatus tests), they were paid $10.

The sample was made up of 73 women and 44 men, $\underline{M}_{age} = 22.5$, $\underline{sd}_{age} = 3.4$, range 18-30 years. For Session 4, examinees were recruited during the experiment (all were told that they could participate in Session 4, but they had to be scheduled for a one-on-one session). Eighty of the examinees completed Session 4. In a few cases, examinees failed to understand the instructions (in the paper & pencil tests), or performed at an unacceptable level (e.g., one tap in the tapping test). Whenever possible, when data were obtained from multiple trials of a test, such data were discarded in favor of an average of 'good' trials. However, in some cases where such imputation was not possible, the respective analysis is based on a reduced sample (noted in the results by the change of at most a few degrees of freedom). Analyses were conducted to ascertain whether any non-representativeness in the examinees who returned for Session 4, compared against those who did not return. No significant differences were found for ability tests, both in the perceptual speed tests or the psychomotor tests. Moreover, analyses were computed for a "constant sample" (examinees who completed the entire four-session experiment), and no significant differences were found. Thus, all of the results reported below are from the maximum number of examinees for each variable.

Apparatus. For the criterion Kanfer-Ackerman Air Traffic Controller Task© (ATC), the keyboard Choice RT and Simple RT tests, instructions, simulation programming and presentation, and response collection were performed with Compaq and IBM 80486 PC computers, with standard keyboards and NEC 4FG display monitors. For paper and pencil tests and tasks, instructions (and timed start-stop directions) were presented over a public address system, using prerecorded minidiscs. Examinees were tested in groups of up to 10 at a time, in individual carrels for the computer-based tasks and at separate tables for the paper & pencil tests.

Ability Testing — Reference Tests. The battery of ability tests was developed to assess two broad categories of Perceptual Speed (PS) abilities (those identified most closely with psychomotor abilities): Perceptual Speed-Simple (PS-Simple) ability tests (where the items are all easy, but the examinee must complete as many as possible in a short period of time), and Perceptual Speed-Complex (PS-Complex), where more complex items required search and look-up strategies. An extensive discussion of these factors is provided in Ackerman, Kanfer, & Goff (1995). For PS-Simple, the following tests were used: (1) Number Comparison (the examinee must rapidly decide whether a pair of numbers is identical or different), (2) Name Comparison (a verbal analogue to the number comparison test), and (3) Number Sorting (in each item, the examinee must decide which of 5 numbers is the largest). For each test, 5 parts lasting 1.5 min/part were given. For PS-Complex, the following tests were used: (1) Directional Headings (in each item, the examinee must review an arrow, a heading in degrees, and a compass abbreviation, and decide if they are concordant or discrepant); (2) Dial Reading (in each item, the examinee must choose the

correct reading for an analog dial -- that sometimes must be interpolated); and (3) Table Reading (for each item, the examinee must locate a single entry in a large table of numbers, after locating the row and column coordinates). The Directional Headings test had two parts (6 min total testing time), and the other two tests had only one part (8 min for the Dial Reading test and 6 min for the Table Reading Test).

Choice/Simple Reaction Time Tests -- Keyboard.

1. Nine-Choice Reaction Time (RT). Stimuli were digits 1-9 (Although an 8-choice RT test may have been preferred for comparison purposes, this choice was dictated by the layout of the numeric keyboard on the standard computer keyboard). Responses were made using the same number keys on the computer numeric keypad.
2. Four-Choice RT. Stimuli were digits 1,2,4,5. Responses were made using the same number keys on the computer numeric keypad.
3. Two-Choice RT. Stimuli were digits 1,2. Responses were made using the same number keys on the computer numeric keypad.
4. Simple RT. Stimulus was the digit 1. Responses were made using the same number key on the computer numeric keypad.

For all the choice RT tests, each trial consisted of a focus dot for 800 msec, the stimulus presentation, and feedback (RT, Average RT, and cumulative accuracy over a block of trials). One block = 25 trials. Performance was measured as the mean RT in msec for correct responses. The Choice-RT tests had stimulus *uncertainty* and temporal *certainty*. For the Simple RT task, a random duration focus dot was used to introduce time uncertainty, given the lack of stimulus uncertainty. Thus, the Simple RT task had stimulus *certainty* and temporal *uncertainty*. The focus dot was displayed for durations with a boundary of 800 to 1200 msec.

Touchpanel Tests - Hardware. The touchpanel tests were implemented on a system which consisted of 15" computer monitors, and factory-installed touchpanel overlays (using analog capacitive technology). The main advantage of this system, for research purposes, is that these systems allowed for both finger input and for a stylus input (the Microtouch TouchPen® is about the size of a pencil, and is tethered to the back of the monitor). The dual-input aspect of these systems allows for evaluation of sources of examinee interaction for method variance and reliability. The touchpanel monitors provide x,y position input (similar to that provided by a standard computer mouse), when the finger or TouchPen actually touches the monitor surface. Input is obtained through a serial port connection, at 9600 baud (which translates roughly to 192 data samples/second). According to the manufacturer, the monitors are capable of responding to touch within 3 ms. In addition, the Mitsubishi monitors are capable of displaying a resolution of 1152 Horizontal x 864 Vertical pixels. The touch position input is then scaled to the screen resolution, yielding nearly one million identifiable unique input locations on the screen. All instructions, stimulus presentation and response collection were performed on DELL or IBM Pentium 90 or 100 MHz computers, with audio presented through the internal computer speaker or external headphones with SoundBlaster hardware, and Microtouch/Mitsubishi monitors, with either

the TouchPen or index finger (preferred hand) as input.

Based on prior pilot testing, it was found that with extensive testing (generally longer than 10-15 min), examinees reported that it was uncomfortable to hold their arms extended nearly straight-out to make optimal contact with the touchpanel. A rather low-technology solution was found, which was to insert a closed 3" three-ring binder under the monitor (with the largest part of the wedge under the front of the monitor). This has the effect of tilting the monitor up, so that it presents an oblique surface to the examinee. Glare on the monitor surface was minimized by using floor-standing incandescent lights (so no direct light shone on the monitors). The examinees were also elevated in sitting position by placing them on stools (at table height -- roughly 30"), instead of chairs. As a result, examinees could comfortably see the monitors and hold the stylus or the index finger downward to the monitor.

Touchpanel Tests - Software Platform. The software platform was developed by Pearson Technical Software, Inc. for Microsoft Windows 3.1. For Experiment 1, the software platform included only tests with discrete responses. (In Experiment 2, the platform was expanded to allow for continuous response tasks -- namely maze tracing and mirror tracing.)

Feedback/Knowledge of Results. Audio (WAV format) files were created that indicated correct or incorrect responses, and then played during the task. For example, when the examinee had a successful 'tap' on the target square of the tapping task, an auditory 'beep' was heard, and when the examinee made an error (tapping outside of the target square) an auditory 'buzz' was heard.

Finally, the computerized instruction platform provided both graphical display and auditory presentation of task instructions (where the displays are bitmapped graphics files and the auditory sequences are WAV files). As such, the examinee required minimal interaction with the examiner during the testing sequence, from initial instruction to final testing.

Touchpanel Tests - Tasks. The displays for the individual psychomotor tests in this sequence are illustrated in **Figure 1**.

1. Single Tapping. Examinee was presented with a single target square and was instructed to tap it as rapidly as possible with either a TouchPen or finger, depending upon the task condition.

2. Alternate Tapping. Examinee was presented with two target squares and was instructed to tap them as rapidly as possible in alternating order with either a TouchPen or finger, depending upon the task condition.

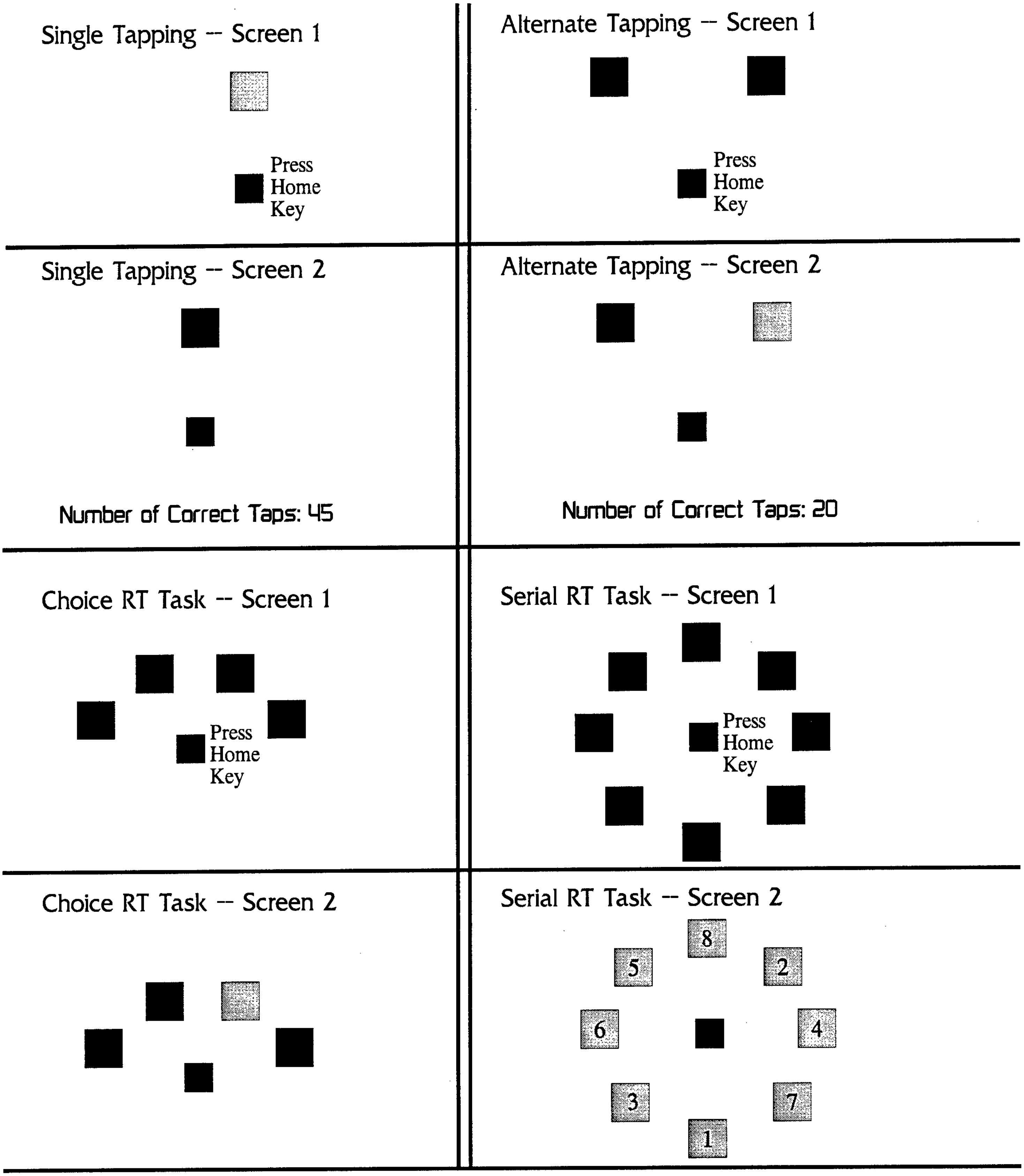


Figure 1. Configural display of computer screens for Choice RT, Serial RT, Single Tapping and Alternate Tapping psychomotor tests. The home key is always presented in red color until the participant touches and holds the key (then it changes to black). In the Choice RT, Simple RT, and Tapping tests, the correct stimulus to-be-touched changes from purple to red after the wait period. In the Serial RT test, all of the colored squares are replaced with numbered squares after the wait period.

All tapping tasks began with a variable 1000 to 2000 msec hold on a home key, followed by a change in target square color signaling trial commencement, then feedback (number of correct taps per trial, number of error taps per trial). All target squares were of equal size, 2.4 x 2.4 cm. Performance was measured as number of correct taps within the trial time limit of 15 seconds. One block of 5 trials was administered for each test.

Choice RT. Four different versions of the Choice RT test were created, as follows.

3. Eight-choice RT. Examinee was presented with 8 squares of equal size, equally distant from the home key. Responses were made by touching the correct target either with a TouchPen or finger, depending upon the task condition.

4. Four-choice RT. Same as 8-choice, with only 4 squares of equal size.

5. Two-choice RT. Same as 8-choice, with only 2 squares of equal size.

6. Simple RT. Same as 8-choice, with only 1 square.

All touchpanel choice RT tasks began with a variable 400 to 800 ms hold on a home key, followed by a change in target square color (analogous to stimulus presentation), then feedback (Trial RT, average RT [for correct trials only], and cumulative accuracy over a block of trials). All target squares were of equal size, 2.4 x 2.4 cm. One block = 25 trials. Each test consisted of two blocks of trials (total = 50 trials).

Performance, measured as mean RT in milliseconds, was calculated for correct responses only. All multiple-choice conditions presented the examinee with an uncertain stimulus after a variable temporal delay. All simple RT conditions presented the examinee with a certain stimulus after a variable temporal delay.

Serial RT. In the Serial RT paradigm, the examinee was instructed to press all of the stimulus squares in numerical order.

7. Eight-item Serial RT. Examinee was presented with 8 target squares arranged equidistant from the "home key" in a circular pattern.

8. Four-item Serial RT. Same as 8-item Serial RT, with only 4 squares of equal size.

All Serial RT tasks began with a variable 400 to 800 msec hold on a home key, followed by the random numbering of all target squares, then feedback (same as the Choice RT feedback). All target squares were of equal size, 2.4 x 2.4 cm. One block equaled 25 trials. Three blocks of trials were administered for each test. Performance, measured as mean "RT" (which is actually a total completion time) in ms, was calculated for correct responses only.

Apparatus Tests. For the apparatus tests in Session 4, the following tests were used, all manufactured by Lafayette Instruments. Instructions were presented orally by the experimenter, according to a written script. The apparatus tests are shown in the photographs in **Figure 2**. All timings (total trial time) were accomplished with a digital stopwatch, operated by the experimenter:

1. Finger Tapper. A manual finger board with internal counter (Lafayette Model 32726).

2. Single and Alternate Tapping: Tapping Board (Lafayette Model 32012), with hand-held tethered stylus and electromechanical impulse counter. Tapping board target size was 3.25" square. For alternate tapping, the centers of the targets were separated by 14.375".

3. Maze Tracing. Maze Tracing was assessed using the Trites Maze Coordination Test (Lafayette Model 32731), with hand-held tethered stylus, electromechanical impulse counter (for errors -- counted when the stylus comes into contact with the sides of the maze). Completion time was assessed via stopwatch operated by the experimenter (for Maze completion time).

Kanfer-Ackerman ATC Task©. Details of the ATC task have been provided elsewhere in the literature, and are also presented in **Appendix A** (e.g., Ackerman, 1988, 1990, Ackerman & Kanfer, 1994; Goska & Ackerman, 1996; Kanfer & Ackerman, 1989). This task was chosen because it is procedural, complex, and involves consistent stimulus-response mappings. Extensive data have been collected with this task, including performance from nearly 5,000 participants across nearly 20 different task and participant-sample configurations (Ackerman & Kanfer, 1994), which makes it possible to predict learning and performance characteristics for specific samples and practice conditions. The task also is useful because it is representative of other procedural learning tasks that are initially cognitively demanding but are capable of being well learned with practice. Commonly taught tasks of this nature range from learning to type, to drive a car, use a word processor, and the technical aspects of playing a musical instrument. The ATC task has been validated against higher-fidelity simulation tasks, such as the Terminal Radar Approach Control (TRACON) simulator system. While such simulation tasks are nearly an order of magnitude more complex than the ATC task (e.g., 20+ hours of training to asymptotic performance in comparison to 3 - 6 hours of training to asymptotic performance in the ATC task), the correlation between performance in the two tasks is approximately .43 (Ackerman, 1996). Extensive laboratory studies have shown that the ATC task demands cognitive, perceptual speed and psychomotor abilities over the course of skill-acquisition trials (e.g., see Ackerman & Kanfer, 1994 for a compendium description of such studies).

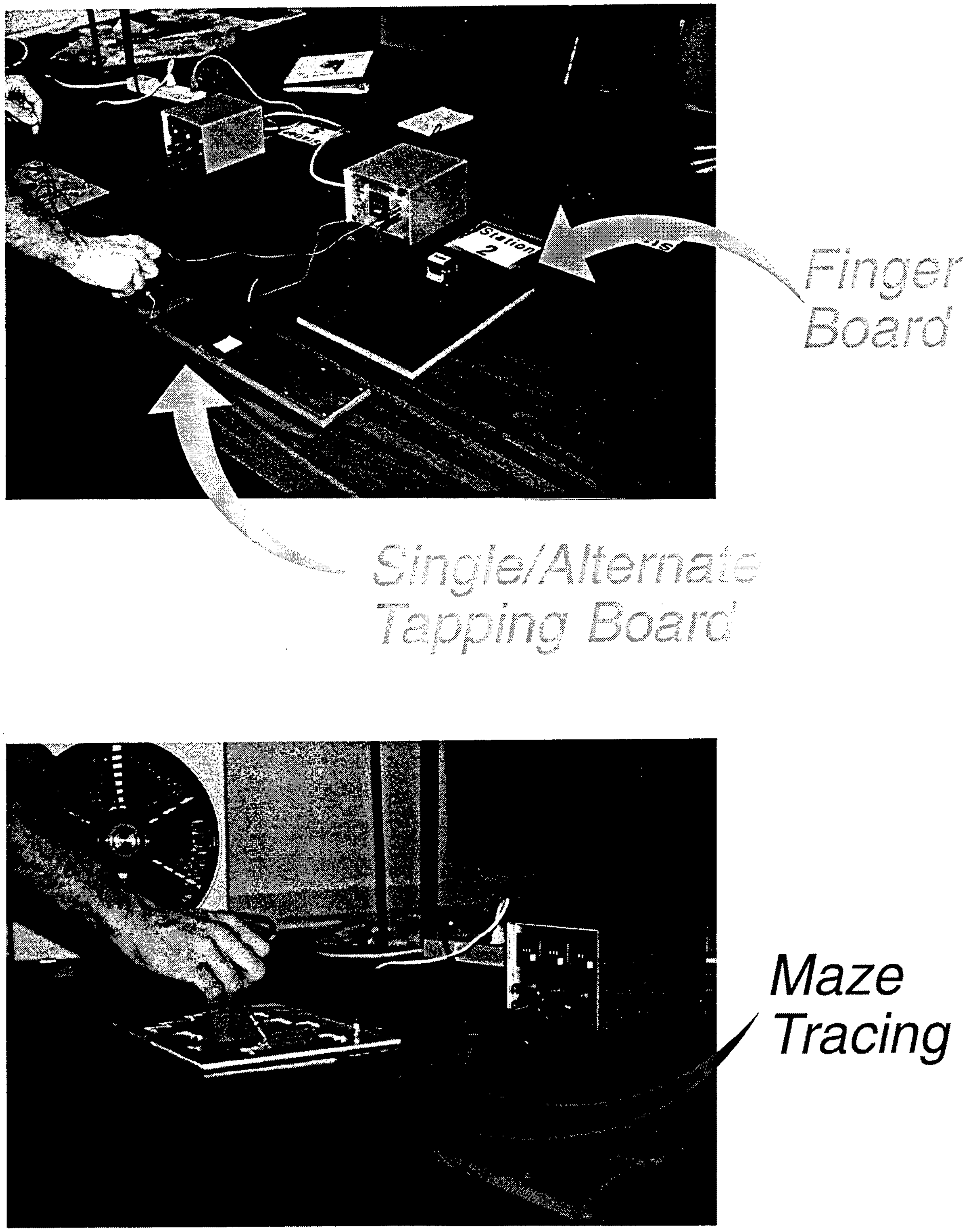


Figure 2. Apparatus tests used in Experiment 1. Shown are the finger board and the tapping board (top photograph), and the Maze Tracing apparatus (bottom photograph), along with the stylus and electromechanical counteres used in the various tests.

Procedure

The procedure for the experiment is illustrated in **Figure 3**. In Session 1, examinees were administered the six paper and pencil Perceptual Speed tests, followed by the keyboard Choice and Simple RT tests. After a break, the following were administered: Part I -- TouchPen input only: Single Tapping (5, 15 sec trials), Alternate Tapping (5, 15 sec trials), Choice/Simple RT (2 blocks of 25 trials each), Serial RT (3 blocks of 25 trials each). After second break, Part II was administered (repetition of all tests, with Finger input only). For each test, three practice trials were administered in the context of the interactive instructions. The practice trials were recorded, but are not analyzed here. Session 2 (which followed Session 1 by two days) included instructions and 12, 10 minute trials of the ATC task. Session 3 (which followed Session 2 by two days) included 6 additional trials of the ATC task, followed by a complete repetition (without instructions/practice trials) of the touchpanel psychomotor tests. Sessions 1-3 were completed in three, 3-hour sessions. At the end of Session 3, examinees were debriefed and asked to fill out a short questionnaire that assessed their attitudes and experiences with the TouchPen and finger input.

Session 4 was administered within a two-week period after Session 1. Examinees completed 5, 15 sec trials of the Finger Board test, 5 15-sec trials of the Single Tapping test, followed by 5 trials of the Maze Tracing test, and then 5 15-sec trials of the Alternate Tapping test. After a break, the entire sequence was immediately repeated. The total time for Session 4 was 45 min. Examinees were separately debriefed for Session 4 at the end of the last apparatus test.

Results

A full analysis of all of the data obtained in this study is beyond the scope of this report (such as analysis of trial-level data, means vs. medians, gender differences, and so on). Instead, several critical issues are reviewed, as follows: (1) Basic psychometric issues (means, reliabilities and short-term practice effects), (2) Comparisons across input formats for Choice/Simple RT, (3) Comparisons across formats for Tapping tests, (4) Construct validity (both within the psychomotor testing paradigm and in relation to measures of perceptual speed abilities), and (5) Criterion-related validity (for the ATC task criterion). Each of these issues is evaluated below.

Basic Psychometric Issues. Several descriptive indices were computed for each of the psychomotor tests, regardless of the form of administration. In each case, data were aggregated across trials and blocks of continuous practice (e.g., 2 blocks of 25 trials for the Choice RT/Simple RT tests, 3 blocks of 25 trials for the Serial RT tests, 5 trials of each of the tapping tests and the apparatus Maze test). **Table 1** shows the means, between-subject standard deviations for both test and re-test administrations, the correlations between test and retest scores, and dependent t-test results (test vs. retest). Reviewing the initial means and sd's for the Choice/Simple RT tests, one characteristic of the tests is clearly visible. That is, in the keyboard version of the tests, RT declines substantially from the 9-Choice RT to the Simple RT tests (mean difference = 521 ms), but in the touchpanel tests, a much more shallow drop in RTs was found (80 ms. for the TouchPen input, and 78 ms. in the Finger

**Session 1**

Paper & Pencil Ability Testing
1. Name Comparison
2. Number Sorting
3. Number Comparison
5. Directional Headings
6. Table Reading
7. Dial Reading

Computer (keyboard)
Choice/Simple RT Tests
1. 9-Choice RT
2. 4 Choice RT
3. 2 Choice RT
4. Simple RT

Computer (touchpanel)
Psychomotor Testing
[TouchPen input]
1. Single Tapping
2. Alternate Tapping
3. 8 Choice RT
4. 4 Choice RT
5. 2 Choice RT
6. Simple RT
7. 8- item Serial RT
8. 4-item Serial RT

[Repeat 1-8 for Finger input]

**Session 2**

Kanfer-Ackerman ATC Task
1. Instructions
2. 12, 10-minute trials

**Session "4"**

Apparatus Psychomotor Testing
1. Finger Board
2. Single Tapping
3. Maze Tracing
4. Alternate Tapping

[Immediate retest of 1-4]

**Session 3**

Kanfer-Ackerman ATC Task
1. 6, 10-minute trials

Computer (keyboard) [RETEST]
Choice/Simple RT Tests
1. 9-Choice RT
2. 4 Choice RT
3. 2 Choice RT
4. Simple RT

Computer (touchpanel)
Psychomotor Testing
[TouchPen input] [RETEST]
1. Single Tapping
2. Alternate Tapping
3. 8 Choice RT
4. 4 Choice RT
5. 2 Choice RT
6. Simple RT
7. 8- item Serial RT
8. 4-item Serial RT

[Repeat 1-8 for Finger input]

Figure 3. The presentation order of tests and re-tests across the four sessions of Experiment 1.

**Table 1. Means, standard deviations for initial testing, retesting; Test-Retest correlations and *t*-tests**

| | $M_{initial}$ | $sd_{initial}$ | $M_{retest}$ | $sd_{retest}$ | $r_{test\text{-}retest}$ | $t$ ($M_{retest}$ - $M_{initial}$) |
|---|---|---|---|---|---|---|
| Keyboard 9-Choice RT | 776.ms | 150 | 695 | 127 | .87 | -11.6** |
| Keyboard 4-Choice RT | 545 | 77 | 528 | 71 | .72 | -3.3** |
| Keyboard 2-Choice RT | 381 | 52 | 377 | 43 | .61 | -1.1 |
| Keyboard Simple RT | 255 | 52 | 254 | 43 | .61 | -0.2 |
| TouchPen 8-Choice RT | 550 | 57 | 513 | 59 | .60 | -7.8** |
| TouchPen 4-Choice RT | 504 | 59 | 481 | 55 | .61 | -5.0** |
| TouchPen 2-Choice RT | 493 | 59 | 477 | 54 | .60 | -3.5** |
| TouchPen Simple RT | 470 | 61 | 463 | 59 | .63 | -1.6 |
| Finger 8-Choice RT | 583 | 66 | 567 | 63 | .72 | -3.7** |
| Finger 4-Choice RT | 540 | 76 | 525 | 69 | .70 | -2.9** |
| Finger 2-Choice RT | 527 | 81 | 516 | 69 | .73 | -2.1* |
| Finger Simple RT | 505 | 80 | 494 | 77 | .74 | -2.0* |
| Pen 8 Serial RT | 4094 | 564 | 3752 | 494 | .84 | -12.0** |
| Pen 4 Serial RT | 1667 | 204 | 1560 | 179 | .70 | -7.6** |
| Finger 8 Serial RT | 3897 | 612 | 3664 | 484 | .85 | -7.7** |
| Finger 4 Serial RT | 1649 | 218 | 1561 | 185 | .70 | -5.9** |
| Finger Board | 77.taps | 13 | 81 | 10 | .69 | -4.1** |
| TouchPen Single Tapping | 93 | 12 | 96 | 12 | .74 | -3.5** |
| Finger Single Tapping | 90 | 11 | 91 | 11 | .84 | -3.0** |
| Apparatus Single Tapping | 105 | 12 | 107 | 13 | .87 | -2.4* |
| TouchPen Alternate Tap | 35 | 6 | 38 | 6 | .59 | -5.89** |
| Finger Alternate Tap | 35 | 6 | 38 | 6 | .53 | -4.76** |
| Apparatus Alt Tap | 55 | 9 | 61 | 10 | .92 | -12.1** |
| Maze Tracing time | 19.sec | 6 | 17 | 6 | .89 | 5.71** |
| Maze Tracing errors | 8.errors | 4 | 7 | 5 | .84 | 3.2** |

Note: *$p < .05$; **$p < .01$

input conditions). On reflection, the reason for this seemed clear -- the overwhelming stimulus-response compatibility in the touchpanel tests appears to substantially influence total RT, almost regardless of the number of choices. (That is, each of the touchpanel Choice and Simple RT tests require that the examinee place the TouchPen or his/her finger on the only green stimulus that appears on the monitor, regardless of the number of other choices, a very direct response.) However shallow the decline in RTs with reduction of number of stimulus choices, though, the decline is generally orderly and consistent across the two forms of response input.

A second aspect of the data that appears striking is that when the Single and Alternate Tapping means are compared, the Apparatus test performance was substantially better, on average, than the performance on the touchpanel tests (means were over $1\sigma$ higher in both he Single and Alternate Tapping tests). The reason for these difference across tests is less clear. One explanation is simply the result of larger stimuli -- the tapping squares in the apparatus version are 3.25" square, and they are 1" square in the touchpanel version. This explanation would indicate that the differences in performance are mainly a function of Fitt's law, relating the size of the stimulus target to response time for tapping.

Otherwise, the data across all forms of input and administration show two important characteristics: First, all but the keyboard Two-choice RT and Simple RT tests showed significant improvement in performance from test to re-test. When put into the context of mean performance changes, keyboard Choice and Simple RTs improved from .02 to $.54\sigma$, the Pen and Finger Choice RT/Simple RTs improved from .12 to $.65\sigma$, the Serial RTs from .38 to $.61\sigma$, the Tapping tests improved from .09 to $.67\sigma$, and the Maze tracing improved $.33\sigma$ in completion speed and $.25\sigma$ in reduction of errors. Second, test-retest reliabilities (after 4 intervening days in the computerized tests, and after a 5-minute break in the apparatus tests) were all quite satisfactory for short-duration tests (e.g., the tapping tests only represent 75 sec of behavior, 5 15-sec trials).

Three considerations suggest that these results provide very encouraging results, namely: The Spearman-Brown Prophecy formula (which indicates how a test reliability increases as a function of the length of the test); the wide-ranging literature on the Power Law of Practice for individual performance (that indicates how increasing practice yields greater stability of performance -- e.g., see the review by Newell & Rosenbloom, 1981); and a few studies in the differential domain that suggest increasing stability of individual differences (e.g., Reynolds, 1952a, 1952b; see Ackerman, 1987 for a review). In light of the fact that the obtained reliabilities ranged from .53 to .92, it seems that by extending either the practice on these tests prior to assessment, or increasing the number of trials administered (or both), test reliabilities for these psychomotor tests that are nearly as high as the reliabilities of longer traditional paper and pencil tests of cognitive and perceptual speed abilities can be achieved. Nonetheless, in an operational testing environment, one must be careful to avoid situations where examinees bring vastly different practice experiences to the testing situation (e.g., for an empirical demonstration and discussion of differential practice in the spatial ability domain, see Ackerman & Lohman, 1990).

Cross-comparisons: Choice RT and Simple RT. **Table 2** shows the intercorrelations among the three different initial administrations of the Choice RT and Simple RT tests.[1] The keyboard results are generally concordant with the prior literature (e.g., see Ackerman, 1990), in that each of the within-task correlations show a simplex-like pattern, as the task decreases in complexity (from 9 to 4 to 2 Choice RT to Simple RT). On the other hand, the relatively higher correlations among the touchpanel tests (higher in fact than the respective test-retest reliabilities), and the similar mean RTs for the tests, support the observation that the high stimulus-response compatibility among the various choice and simple RT tests yielded essentially equivalent data (both in terms of average performance and relative rank-ordering of individuals). As such, the four TouchPen and Finger Choice/Simple RT tests were collapsed into single composites in further discussion.

Cross-comparisons: Tapping and Alternate Tapping. For the tapping tests, the formats of administration included both apparatus and computer touchpanel. The respective cross-test intercorrelations (and test-retest correlations) are shown in **Table 3.** In some sense, this table can be thought of as representing a multi-trait (Single Tapping and Alternate Tapping), multi-method (Pen, Finger, Apparatus) matrix. With this perspective, the following observations can be made: (1) Cross-method correlations are higher for the Single Tapping tests than for the Alternate Tapping tests; (2) More commonality is found for same trait, multiple methods (e.g., Single Tapping-Finger with Single Tapping-Apparatus [$r$ = .63] as compared with multiple trait, single method -- Single Tapping-Finger, Alternate Tapping-Finger [$r$ = .38]), though there is much common variance across all tests (with the highest commonality found with the single tapping, a moderate amount of common variance for the alternate tapping tests, and the least common variance with the finger board). Although this table demonstrates both cross-test commonality and substantial commonality for both test-retest same-form, and test-retest alternative format, the relations among all of these tests (along with the choice/simple RT tests) are hard to summarize, based on simple correlations. The next section describes a factor analysis of the psychomotor tests, in order to provide a summary of intercorrelations among the tests in the construct space.

---

[1] To avoid inevitable confusion, throughout this report, whenever correlations between different measures are computed, scores were reflected as needed (by multiplying the scores by -1.0), so that a positive correlation between any two measures means that good performance in the first measure is positively related to good performance in the second measure. For example, a positive correlation between a Choice RT test (where small numbers mean a short reaction time, and thus good performance) and a paper and pencil test (where high scores mean more correct answers, and thus good performance), means that examinees who performed well on the Choice RT test were more likely to perform above average on the paper and pencil test.

Table 2. Cross-Correlations: Choice and Simple RT Tests

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Keyboard 9 Choice RT | | | | | | | | | | | |
| 2. Keyboard 4 Choice RT | **.74** | | | | | | | | | | |
| 3. Keyboard 2 Choice RT | **.49** | **.67** | | | | | | | | | |
| 4. Keyboard Simple RT | **.38** | **.59** | **.75** | | | | | | | | |
| 5. TouchPen 8 Choice RT | .40 | .52 | .45 | .53 | | | | | | | |
| 6. TouchPen 4 Choice RT | .42 | .57 | .48 | .56 | **.80** | | | | | | |
| 7. TouchPen 2 Choice RT | .38 | .51 | .51 | .55 | **.73** | **.88** | | | | | |
| 8. TouchPen Simple RT | .38 | .50 | .46 | .56 | **.69** | **.81** | **.84** | | | | |
| 9. Finger 8 Choice RT | .39 | .46 | .50 | .51 | .63 | .73 | .72 | .71 | | | |
| 10. Finger 4 Choice RT | .33 | .42 | .46 | .48 | .58 | .66 | .68 | .68 | **.86** | | |
| 11. Finger 2 Choice RT | .35 | .45 | .44 | .49 | .56 | .65 | .65 | .65 | **.83** | **.94** | |
| 12. Finger Simple RT | .42 | .52 | .48 | .52 | .56 | .66 | .65 | .64 | **.75** | **.85** | **.87** |

---

**Bold** = same method, different trait

Underline = same trait, different method

All correlations significant, $p < .05$ (df = 115).

Table 3. Cross-Correlations: Tapping and Alternate Tapping

| Initial Testing | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. TouchPen Single Tap | | | | | | | | | | | | | |
| 2. Finger Single Tap | **.69** | | | | | | | | | | | | |
| 3. Apparatus Single | **.58** | **.63** | | | | | | | | | | | |
| 4. TouchPen Alternate Tap | *.38* | .15 | .02 | | | | | | | | | | |
| 5. Finger Alternate Tap | .30 | *.38* | .06 | **.50** | | | | | | | | | |
| 6. Apparatus Alt Tap | .47 | .36 | *.40* | **.43** | **.24** | | | | | | | | |
| 7. Finger Board | .35 | .20 | .53 | .03 | -.10 | .16 | | | | | | | |
| Retesting | | | | | | | | | | | | | |
| 8. TouchPen Single Tap | .74 | .76 | .66 | .13 | .25 | .40 | .30 | | | | | | |
| 9. Finger Single Tap | .60 | .84 | .70 | .07 | .25 | .42 | .17 | **.84** | | | | | |
| 10. Apparatus Single | .67 | .64 | .87 | .13 | .10 | .53 | .28 | **.71** | **.76** | | | | |
| 11. TouchPen Alternate Tap | .40 | .27 | .19 | .59 | .50 | .44 | .07 | *.42* | .38 | .27 | | | |
| 12. Finger Alternate Tap | .28 | .29 | .27 | .30 | .53 | .34 | .03 | .39 | *.43* | .34 | **.62** | | |
| 13. Apparatus Alt Tap | .37 | .33 | .35 | .39 | .24 | .92 | .10 | .34 | .39 | *.45* | **.43** | **.35** | |
| 14. Finger Board | .48 | .25 | .44 | .20 | -.05 | .37 | .69 | .37 | .28 | .43 | .18 | .11 | .27 |

**Bold = same trait, different method**; *Italics = different trait, same method*; Double-underline, test-retest reliability

Construct Validity. Composite scores for the three Choice/Simple RT test formats and the two Serial RT test formats were first formed (by adding unit-weighted *z*-scores for each set of tests). All test measures were derived from the initial testing session. This amalgamation yielded 12 variables (Keyboard, TouchPen, and Finger Choice/Simple RT; TouchPen and Finger Serial RT composites; Finger Board; TouchPen, Finger, and Apparatus Single Tapping; and TouchPen, Finger, and Apparatus Alternate Tapping). (The Maze Tracing test was excluded from this analysis because it showed essentially zero overlap with any of the other psychomotor tests. Given that none of the other psychomotor tests, either apparatus or touchpanel, involved continuous movement, it seemed most likely that the uniqueness of the Maze Tracing test was mainly a function of the differences in underlying abilities used to perform the task. As such, it can be said that this test demonstrated substantial discriminant validity). A principal factor solution was derived, with squared multiple correlations as initial communality estimates. The Humphreys-Montanelli (1975; Montanelli & Humphreys, 1976) parallel analysis method was used to select the number of factors that underlie the correlation matrix (the analysis yielded a recommendation for 4 factors). The principal axis solution was then rotated to an orthogonal (Varimax) rotation, which is shown in **Table 4**.

Interpretation of the factors is relatively straightforward, as follows: Factor I is defined mainly by salient loadings of the Serial RT and Choice/Simple RT tests, and is thus called "Serial/Choice RT;" Factor 2 is defined mainly by the Single Tapping tests, and is thus called "Single Tapping;" Factor 3 is defined mainly by the Alternate Tapping tests, and is thus called "Alternate Tapping;" and Factor 4 is minimally defined by a singleton loading from the Finger Board. However, because the Finger Board also loads significantly on Factor 2, it may be best to think of Factor 4 more as a construct that is unique to aspects of the Finger Board test, rather than a common factor, per se.

While the broad interpretation of these factors is clear, it is also apparent that common method variance also played a role in defining the common factor structure. The Serial/Choice RT factor (Factor 1) appears to also capture a significant amount of "Finger" method variance, the Single Tapping factor (Factor 2) appears to capture some "Apparatus" method variance, and the Alternate Tapping factor (Factor 3) appears to capture the remaining "TouchPen" method variance. More elaborate multivariate procedures could probably be used to tease apart the actual contribution of each method to the total variance, but that kind of analysis is more of academic interest than of practical usefulness.

An additional view of construct validity is provided by examining correlations between the psychomotor tests and the two Perceptual Speed ability composites: PS-Simple and PS-Complex. These correlations are shown in **Table 5**. The Maze Tracing and Finger board scores (both apparatus tests) show no significant overlap with the paper & pencil based composites of PS-Simple or PS-Complex abilities. Otherwise, the remaining psychomotor tests show moderate-to-substantial overlap with the PS abilities, but given the commonality among the two PS abilities, the respective psychomotor-PS correlations tend to be similar. (Only the Keyboard Choice/Simple RT test composite and the Apparatus Single Tapping test had significantly higher correlations with the PS-Complex composite than with the PS-Simple

Table 4. Factor solution (Varimax) for the psychomotor tests and composites.

| | | I | II | III | IV |
|---|---|---|---|---|---|
| 1. | Keyboard Choice RT/Simple RT | **.645** | .258 | .065 | .272 |
| 2. | TouchPen Choice RT/Simple RT | **.787** | .148 | .276 | .136 |
| 3. | Finger Choice RT/Simple RT | **.824** | .163 | .039 | -.045 |
| 4. | TouchPen Serial RT | **.670** | .116 | **.352** | -.028 |
| 5. | Finger Serial RT | **.807** | .169 | .145 | -.104 |
| 6. | Finger Board | -.010 | **.416** | -.003 | **.557** |
| 7. | TouchPen Single Tapping | .224 | **.715** | **.333** | .108 |
| 8. | Finger Single Tapping | **.368** | **.791** | .030 | -.181 |
| 9. | Apparatus Single Tapping | .033 | **.837** | -.027 | .266 |
| 10. | TouchPen Alternate Tapping | .266 | .050 | **.846** | .000 |
| 11. | Finger Alternate Tapping | **.593** | .127 | **.356** | -.243 |
| 12. | Apparatus Alternate Tapping | .069 | **.444** | **.426** | -.009 |

-------

Note: Salient loadings in **bold**

Table 5. Psychomotor test correlations with PS/Simple, PS/Complex, ATC Session 1, ATC Session 6

| | PS Simple | PS Complex | ATC Session 1 | ATC Session6 | ATC Overall |
|---|---|---|---|---|---|
| PS/Simple | | .65** | .57** | .49** | .54** |
| PS/Complex | .65** | | .68** | .60** | .65** |
| ATC Session 1 | .57** | .68** | | .74** | |
| ATC Session 6 | .49** | .60** | .74** | | |
| Keyboard Choice/Simple RT | .44** | .62** | .50** | .46** | .49** |
| TouchPen Choice/Simple RT | .36** | .42** | .38** | .36** | .38** |
| Finger Choice/Simple RT | .38** | .42** | .41** | .42** | .45** |
| TouchPen Serial RT | .51** | .46** | .48** | .41** | .46** |
| Finger Serial RT | .55** | .53** | .55** | .53** | .58** |
| Finger Board | .12ns | .16ns | .21ns | .10ns | .15ns |
| TouchPen Single Tapping | .20* | .32** | .29** | .26** | .30** |
| Finger Single Tapping | .18ns | .28** | .25** | .18ns | .21ns |
| Apparatus Single Tapping | .10ns | .28* | .26* | .21ns | .27* |
| TouchPen Alternate Tap | .34** | .26** | .32** | .26** | .26** |
| Finger Alternate Tap | .45** | .39** | .40** | .37** | .39** |
| Apparatus Alt Tap | .27* | .12ns | .13ns | .15ns | .14ns |
| Maze Tracing time | .14ns | .20ns | .07ns | .03ns | .06ns |
| Maze Tracing errors | .09ns | .01ns | .05ns | .10ns | .07ns |

N*ote*: $*p < .05$; $**p < .01$; ns = not significant

composite [$t(114) = 2.93$, $p < .01$, and $t(77) = -1.98$, $p < .05$], for keyboard Choice/Simple RT and Apparatus Single Tapping, respectively.) Such results indicate that there is indeed both common and unique variance between the PS ability composites and the psychomotor tests -- concordant with previous analysis of these constructs (e.g., see discussion by Ackerman, 1990).

Criterion-related validity. The Kanfer-Ackerman ATC task is probably not an ideal criterion task (partly because it is initially somewhat complex, and partly because asymptotic performance is typically not attained until five or six hours of practice). However, for the initial validation purposes of the current experiment, it provides a good backdrop for examination of psychomotor test predictive validity -- especially because extensive ability-performance data already exist for this task (see Ackerman, 1988, 1990; Ackerman & Kanfer, 1994; Goska & Ackerman, 1996; Kanfer and Ackerman, 1989). **Table 5** also shows the correlations between the PS composites, psychomotor tests, and performance on the ATC criterion task for the first 30-min session, the final (sixth) 30-min session, and overall performance (average across all six practice sessions).

As expected from the prior literature, performance across the six sessions of ATC practice is very well predicted by the two PS ability composites (with higher correlations from the PS-Complex ability). Also, there is substantial commonality between Session 1 ATC performance and Session 6 ATC performance ($r = .74$). Nonetheless, many of the psychomotor tests showed positive, significant, and substantial correlations with the criterion task performance. Highest correlations were found for the Serial RT tests, followed by the Choice/Simple RT tests, and then the Alternate and Single Tapping tests. Neither the Finger Board nor the Maze Tracing tests had validity coefficients significantly greater than zero. Moreover, in accordance with the historical literature that show more stable validities for psychomotor predictors over training (Brown & Ghiselli, 1952), and Ackerman's (1988) theory of ability-performance relations during skill acquisition, the psychomotor predictors showed relatively small declines in validity coefficients from ATC Session 1 to Session 6.

Raw correlations, though, do not allow one to take account of common variance across predictor variables. To provide a meaningful comparison between the various predictors (paper & pencil, computerized psychomotor, and apparatus psychomotor), a series of hierarchical multiple regression/correlation analyses was performed. These results are shown in **Table 6**. Prediction equations were derived for two different sets of predictors: the first with only apparatus psychomotor tests and the other non-psychomotor predictors, the second with only touchpanel psychomotor tests and the other non-psychomotor predictors. Both Session 1 and Session 6 ATC task performance was predicted (though ATC Session 1 performance was also entered into the prediction equation for Session 6 ATC performance). The results are striking. The apparatus psychomotor tests, by themselves, accounted for only 8% of Session 1 ATC performance, and 5% of Session 6 ATC performance. In contrast, the touchpanel psychomotor tests accounted for 32% of Session 1 ATC performance, and 27% of Session 6 performance. Thus, while not entirely equivalent, a substantial degree of validity was found for the touchpanel tests in the aggregate.

Table 6. Hierarchical Multiple Regression/Correlation Results: <u>Prediction of ATC Performance</u>

| | Session 1 | | | Session 6 | | |
|---|---|---|---|---|---|---|
| Step | $R^2$ | Increment in $R^2$ | $F$ to add | $R^2$ | Increment in $R^2$ | $F$ to add |
| 1. Apparatus Tests[a] | .078 | .078 | 1.58(4,74)ns | .050 | .050 | .98(4,74)ns |
| 2. Perceptual Speed/Simple | .377 | .299 | 8.84(1,73)** | .274 | .224 | 22.50(1,73)** |
| 3. Perceptual Speed/Complex | .512 | .135 | 19.94(1,72)** | .387 | .113 | 13.34(1,72)** |
| 4. Session 1 ATC Performance | .NA | .NA | | .569 | .181 | 29.86(1,71)** |
| **Total** | **.512** | **.NA** | **12.60(6,72)**** | **.569** | **.NA** | **13.38(7,71)**** |

| | Session 1 | | | Session 6 | | |
|---|---|---|---|---|---|---|
| Step | $R^2$ | Increment in $R^2$ | $F$ to add | $R^2$ | Increment in $R^2$ | $F$ to add |
| 1. Touchpanel Tests[b] | .322 | .322 | 13.04(4,110)** | .266 | .266 | 9.98(4,110)** |
| 2. Perceptual Speed/Simple | .409 | .087 | 16.08(1,109)** | .323 | .057 | 9.18(1,109)** |
| 3. Perceptual Speed/Complex | .530 | .121 | 27.88(1,108)** | .410 | .087 | 15.85(1,108)** |
| 4. Session 1 ATC Performance | .NA | .NA | | .568 | .158 | 39.08(1,107)** |
| **Total** | **.530** | **.NA** | **20.31(6,108)**** | **.569** | **.NA** | **20.08(7,107)**** |

Note: $*p < .05$; $**p < .01$
[a]Apparatus tests (Finger board, Single Tapping, Alternate Tapping, Maze completion time)
[b]TouchPanel tests (Choice/Simple RT, Serial RT, Single Tapping, Alternate Tapping)
ATC = Performance on the Kanfer-Ackerman ATC task (mean number of plans landed per trial).

Entering the two PS ability composites, though, indicates the role of common variance between the psychomotor and PS tests. For Session 1 ATC performance, the total variance accounted for was essentially identical, whether one used the apparatus tests in conjunction with the PS tests, or the touchpanel tests in conjunction with the PS tests (51% and 53% respectively of the ATC performance variance). With the addition of Session 1 ATC performance, the amount of Session 6 performance variance was identical across the two analysis strategies (57% of the variance). Note that this is an increase in variance accounted for, even as ATC task performance variability declines as the examinees become more skilled with the ATC task (Session 1 $\underline{sd} = 9.87$, Session 6 $\underline{sd} = 8.13$, a reduction of 18% of total between-subjects' variance).

TouchPen vs. Finger input. At the end of the study, examinees were asked to complete a short questionnaire regarding their attitudes and experience with the touchpanel monitors. The first two questions asked about the 'difficulty' of working with the TouchPen and Finger responses, respectively (each question was presented in a 5-point Likert-type scale). The second set of questions asked about situations where the examinee may have thought that he/she had made a good contact with the touchpanel, but the computer may not have given credit for the response (again for TouchPen and for finger input). These questions were presented with an 8-point scale from "never" to "constantly." Three additional open-ended questions asked for comments regarding likes and dislikes in interacting with the touchpanel monitors. One hundred of the 117 examinees completed the questionnaire. For the first pair of questions, the TouchPen was rated "Moderately easy" ($\underline{M} = 2.17$, $\underline{sd} = .91$), and the Finger input was rated between Moderately easy and Neither easy nor Difficult ($\underline{M} = 2.51$, $\underline{sd} = 1.04$). A dependent $t$-test indicated that the examinees found the TouchPen relatively easier than the Finger input -- $t(99) = 2.85$, $p < .01$. For the questions about 'missed' contacts by the computer, both formats resulted in quite similar responses ($\underline{M} = 3.77$ and 3.99, $\underline{sd} = 1.05$, 1.16 for the TouchPen and Finger input, respectively). These values correspond to the "Several Times" response option that the examinees felt that the computer failed to register their responses -- quite modest in the context of the fact that the examinees made an average of 6,500 contacts each (excluding practice trials) with the computer over the course of the entire experiment.

## Experiment 1 Discussion

This experiment demonstrated that the computerized touchpanel system adopted and the software platform developed for assessment of a small set of psychomotor abilities yielded acceptable results, in terms of reliability, construct validity, and criterion-related validity. In several cases, the reliability and validity indices for touchpanel tests rivaled or exceeded those of analogous apparatus tests.

Moreover, though not presented as results, it was also demonstrated that the use of computers and touchpanel monitors removed the four major obstacles to assessment of psychomotor abilities outlined by Melton (1947), and discussed in the introduction to this report, namely: Fabrication costs, Calibration requirements, Examiner training, and

Examiner-to-examinee ratio. For fabrication costs, it was found that (after investment in the monitors and the software platform development), the cost of changing the test design (such as the number of alternative stimuli, the size of stimuli, the distance between targets) was essentially negligible (since these changes could be made with about 5 minutes invested in editing an ASCII text file with a text editor). As for calibration, each of the 10 monitors was calibrated at the beginning of the study (when the monitors were positioned at an oblique angle to the work surface), and showed no need for re-calibration during the study. Examiners only required training in instructing the examinees not to press hard with the TouchPens (and thus relatively untrained undergraduate research assistants were quite capable of supervising the touchpanel testing). Moreover, the computer provided all of the other examinee instruction and training in using the touchpanel systems. Research assistants were on-hand to answer infrequent questions, typically when the examinee didn't pay attention to the instructions on the computer. (In contrast, the apparatus tests required specialized training of the staff in the use of the counters, stopclocks, stopwatches, and script reading.) Further, the fourth problem of low examiner-to-examinee ratio appears to have been solved, in that a full laboratory (10 workstations) required only a single examiner. The ultimate limit on examiner-to-examinee ratio, at least in terms of having examiners readily available for any questions or problems that arise, is probably in the neighborhood of one examiner for every 15 examinees. Finally, it is illustrative to note that while the limited apparatus testing required one new electrical stylus (to replace a broken metal tip) and an electromechanical impulse counter (which died for unknown reasons), no maintenance repairs for the computerized touchpanel tests were required during the experiment.

All in all, this experiment has shown substantial promise for assessment of some psychomotor abilities via computerized touchpanel monitors. One important caution should be noted, and that is that the battery of predictor tests was limited to perceptual speed and psychomotor measures. Ultimate utility of these new measures will depend in part on how much independent incremental variance the new measures have, in the context of cognitive ability measures (for a discussion of such issues, see the meta-analysis by Levine, et al., 1996; and the recent study by Wolfe, 1997)

The kinds of tests that were developed for the first experiment represented the least demanding system (namely, the assessment of discrete responses). The next challenge was to develop a measure of direct continuous-response psychomotor ability, and measure of displaced continuous-response psychomotor ability -- two kinds of psychomotor tests that have received substantial historical empirical investigation. Experiment 2 was designed to assess these new tests of maze tracing and mirror tracing, respectively, and to put the entire set of computerized touchpanel tests into a broader construct space -- specifically to place the psychomotor tests in an ability space defined by broad cognitive abilities and perceptual speed abilities.

## IV. Experiment 2

Overview

Experiment 2 was designed with two primary goals and one secondary goal. The first primary goal was to evaluate the efficacy of two continuous-response psychomotor tests -- maze tracing and mirror tracing. The second primary goal was to evaluate the construct validity of the entire battery of touchpanel psychomotor tests (those from Experiment 1 and the two new tests) in the context of a broad array of cognitive and perceptual speed abilities. The secondary goal was to replicate the normative results regarding the touchpanel psychomotor tests that were evaluated in Experiment 1 (for additional details, see Cianciolo, 1997).

The development of the two new psychomotor tests is described first. Next the cognitive and perceptual speed ability battery is described, followed by the description of the methods involved in Experiment 2 (with differences from Experiment 1 noted in particular).

Development of Continuous-Response Psychomotor Tests

The development challenge for creating continuous-response tests was to evaluate the critical components of the apparatus tests, and attempt to capture as much of the relevant characteristics of the tests in adaptation to the touchpanel methodology. Continuous-response apparatus tests often provide a different kind of direct feedback than that provided in the discrete response apparatus tests.

Maze Tracing. In the traditional maze tracing apparatus (e.g., the Lafayette Instruments Model #32731 -- shown in Figure 2), the examinee is asked to trace along the path of a maze with a stylus as rapidly as possible, while avoiding making errors. The track that the examinee traces is actually inset into the apparatus, and the sides of the maze are raised, so that when the examinee's stylus touches the side of the maze an electrical contact is made (and the error counter increments). In contrast, the touchpanel is a flat surface -- and thus it is not possible to provide direct tactile feedback along with the sound of the error counter. Another concern that was raised in the consideration of such tasks, was that in the one-on-one apparatus test, examinees were rarely (if ever) tempted to simply raise the stylus and move to the end of the maze, without completing the task in the instructed fashion.

The design of the task is shown in **Figure 4**. Feedback, when the examinee crosses the outside edges of the maze, was provided by a loud auditory buzzer, which, in contrast to the apparatus test, continued with 3-sec intervals until the examinee returned to trace inside the maze. The examinee's progress was shown by superimposing square "blocks" over the continuous maze, and changing the color of each completed block from red to black as the examinee's TouchPen or finger made contact. Forward progress was insured by requiring the examinee to move through the maze in the prescribed order. That is, the examinee could not make "progress" in any manner other than forward movement. After pilot testing, we did allow for a small degree of 'corner-cutting' similar to the apparatus test, such that a examinee could still move forward as long as no more than a single block was missed. That is, the examinee could not skip two contiguous blocks, but would have to return and trace the missing blocks.

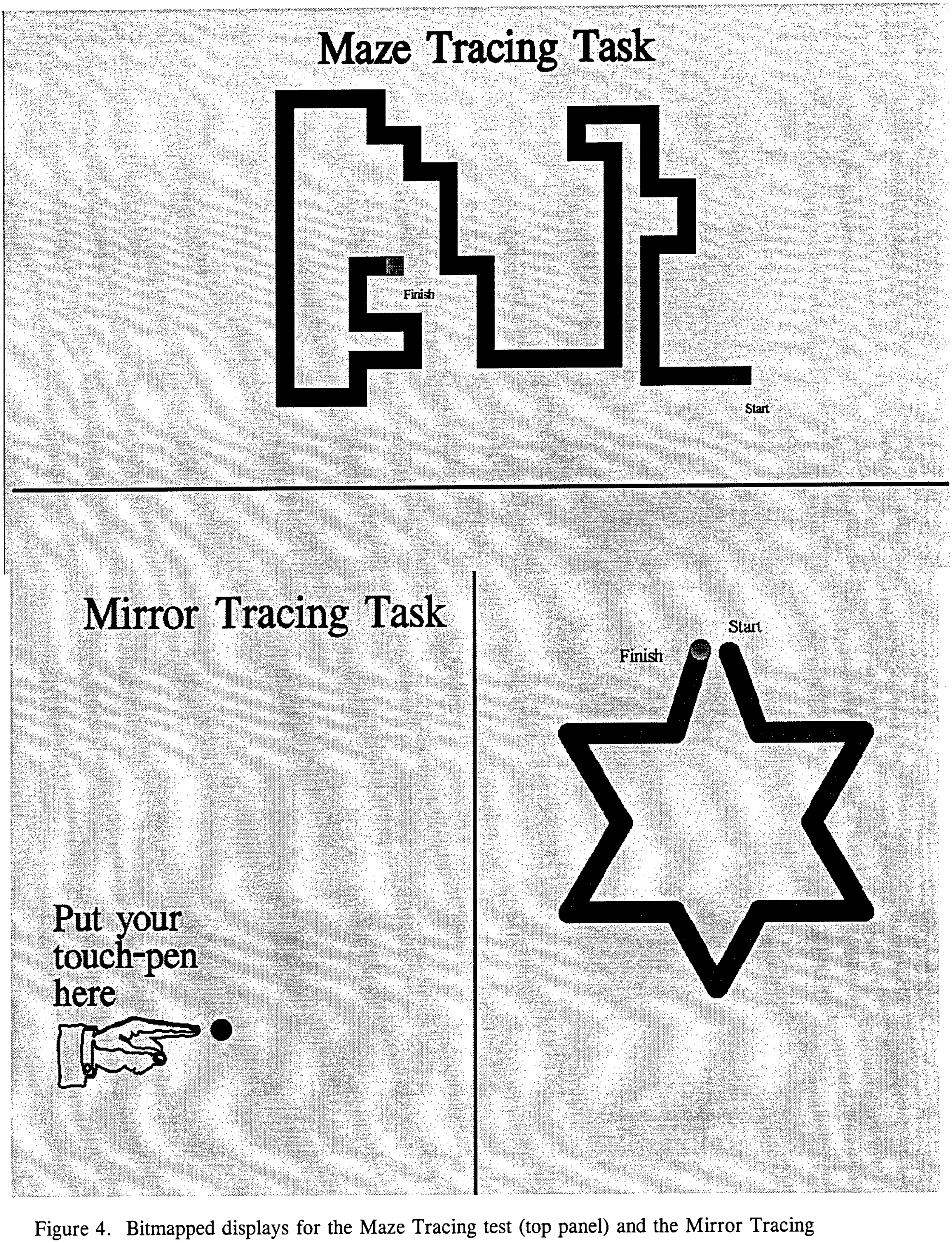


Figure 4. Bitmapped displays for the Maze Tracing test (top panel) and the Mirror Tracing test (bottom panel).

Mirror Tracing. There have been many different instantiations of the mirror tracing task originally described by Snoddy (1920). One prominent apparatus configuration is a largely self-contained system that functions very much like the maze tracing apparatus (with indirect observation through the mirror), in that the track is inset, and the edges of the 'star' provide the metal contacts for the error counter. Typically, a stopwatch is used to assess total completion time. The other prominent apparatus configuration is more like the computer, in that a sheet of paper is provided with a star printed on it. In this version, the examinee draws with a pen directly on the sheet of paper, while observing the paper indirectly through the mirror. Total completion time is measured by stopwatch, but errors are computed afterwards, by physically counting the number of times the pen traced outside of the track. The new computerized instantiation of the test functioned somewhat as a hybrid between the electromechanical version and the paper version. Specifically, no tactile feedback was provided (as with the paper version), but an error tone was sounded whenever the examinee traced outside of the designated track.

However, no mirror was used in the new instantiation of the task. Instead, the examinee traced (with TouchPen or finger) on the left side of the monitor screen (which was blank, other than a 'home key,' with the results of the tracing shown on the right side of the screen, as a continuous line overlaid on the track (see **Figure 4**). The tracing line was displayed in one color (white) when the examinee traced within the track, and another color (purple) when the examinee left the designated track. Examinee progress was also indicated by changing the color of the circles in the track, in a fashion similar to the Maze Tracing test. After a series of pilot tests, it was decided to 'reflect' the output across both the "x" and "y" axes (rather than just the "x" axis, as happens with the mirror tracing apparatus test). That is, as the examinee moved the TouchPen 'up' on the left side of the screen, the path traced moved 'down.' And, as the examinee traced 'left,' the path traced moved 'right' (similar to the procedure used by McDermid & Smith, 1964). Total completion time and errors were displayed after each trail.

Pilot testing suggested that the new test had captured much of the 'feel' of the apparatus version of the mirror tracing test -- examinees were just as likely to get 'stuck' in drawing around the corners of the touch-panel stimulus tracks as they were in the apparatus versions. However, generation of additional tracks, in both the maze tracing and mirror-tracing tasks was a relatively simple task of using an interactive program for drawing line segments that were translated into stimulus mazes.

Ability Test Battery

Ability tests were selected from locally developed and validated batteries (e.g., see Ackerman & Kanfer, 1993; Ackerman & Rolfhus, 1996) [Verbal Analogies, Problem Solving, Math Knowledge, Paper Folding, Verbal Test of Spatial Ability, Spatial Orientation, Finding A's and T's, Finding ∈/¥, Canceling Symbols, Digit/Symbol Substitution, Naming Symbols, Coding, Number Comparison, and Name Comparison], from the Educational Testing Service (ETS) Kit of Factor-Referenced Cognitive Tests [Extended Range Vocabulary and Controlled Associations], or commercial tests [Number Series,

Mechanical Reasoning, Mechanical Knowledge, and Clerical Abilities-2]. Extensive descriptions of these tests can be found in the literature, or by request to the authors.

## Method

### Participants

One hundred nineteen students from an introductory psychology course at the University of Minnesota participated in the study for course credit and $30 cash. All examinees were native English speakers, between 18-30 years old, and had normal or corrected-to-normal hearing, vision, and motor coordination. Two examinees were eliminated from the study, the first due to failure to meet the age requirement, the second due to general failure to follow instructions, leaving a total of 117 examinees for the data analysis. The final sample was made up of 81 women and 36 men, $\underline{M}_{age} = 18.8$, $\underline{sd}_{age} = .72$, range 18-21 years (thus a significantly younger sample, $t\ (232) = 11.47$, $p < .01$, and a slightly higher proportion of women -- 62% in Experiment 1, and 69% in Experiment 2).

### Apparatus

Pencil-and-paper testing during Session 1 was administered in a large classroom with prerecorded instructions and directions presented over a public address system. Up to 40 examinees were tested at a time. During Session 2, pencil-and-paper testing was administered in the same manner, but in a smaller classroom with groups consisting of no more than 10 examinees. The order and presentation of computerized touchpanel psychomotor tests were administered in a manner identical to that in Experiment 1, with the addition of two continuous-response tests (See **Figure 4** for an illustration of the two tests).

Maze Tracing. Mazes were constructed using the same number of turns and proportional segment lengths as the original Trites Maze Coordination Test maze tracing apparatus task (used as an apparatus test in Experiment 1). All Maze Tracing trials began with a variable 500 to 1000 msec hold on a home key, followed by an auditory "ready, set, go" signal and change in home key color signaling trial commencement. After the trial, examinees were provided with feedback for 3 sec. (completion time, average completion time, number of errors). Track width in the TouchPen version and the finger input version were 6mm and 8mm, respectively. Performance was measured as completion time in ms and number of errors. Eight different mazes were created, four for the TouchPen version and four for the finger version.

Mirror Tracing. Patterns were constructed using the same number of corners as the original mirror tracing star (Snoddy, 1920), including a proportional computerized version of the standard star (as the first track to be traced). Examinees traced on the left side of the screen. Examinees monitored their progress by watching their efforts appear in mirrored line transformation within the target pattern on the right side of the screen. Width of the traced line was 1 mm. Width of the pattern segments was 6 mm (Note: because the examinee did not actually place the TouchPen or finger directly on the stimulus, it was not necessary to create a wider track in the Mirror Tracing test). All Mirror Tracing trials followed the same procedure as the Maze Tracing (e.g., home key, "ready, set, go", feedback).

Performance was measured as completion time in ms and number of errors.

Ability Tests. Pencil-and-paper testing included tests to assess the following ability factors:

1. Verbal: Verbal Analogies, Extended Range Vocabulary, Controlled Associations
2. Spatial: Paper Folding, Verbal Test of Spatial Ability, Spatial Orientation
3. Mathematical: Problem Solving, Mathematical Knowledge, Number Series
4. Mechanical: Mechanical Reasoning, Mechanical Knowledge
5. Perceptual Speed-Scanning: Number Comparison, Clerical Abilities-2, Name Comparison
6. Perceptual Speed-Pattern Recognition: Finding A/T, Canceling Symbols, Finding ∈/¥
7. Perceptual Speed-Memory: Naming Symbols, Coding, and Digit Symbol

Procedure

The study took place over three sessions, totaling 9 hours. (See **Figure 5** for a description of the order of presentation.) Session 1 consisted entirely of pencil-and-paper test administration. Session 2 consisted of 1.5 hours of pencil-and-paper testing and 1.5 hours of computerized psychomotor testing through both keyboard and TouchPen (only) input. For Session 3, examinees completed the only the finger input conditions of the touchpanel psychomotor tasks, following an order of presentation identical to Session 2. (The 3rd session also included assessment of personality constructs for an unrelated study.) Two 5-minute breaks were provided in each session. All examinees completed all tasks in the same fixed order.

Results

Because there is much overlap in the tests administered in Experiments 1 and 2, the results will be presented in a way that minimizes redundancy, in the following fashion: The first part of the results section will be devoted to a comparison of results between Experiment 1 and 2, and any differences noted. The second part will present basic descriptive statistics from the continuous-response psychomotor tests. The third part of the results will focus on the evaluation of the entire set of psychomotor tests in the context of the larger ability construct space.

Experiment 1 vs. Experiment 2. In Experiment 2, all of the psychomotor tests that were administered in Experiment 1 were re-administered (except without re-testing, and the TouchPen and Finger response versions of the tests were administered on different days in Experiment 2, rather than in the same session, as in Experiment 1). In addition, the participant pool for Experiment 1 was entirely made up of paid examinees (having been run during the summer months, when the large introductory psychology course participant pool was unavailable). In contrast, the examinee pool for Experiment 2 was entirely made up of Introductory Psychology students, participating for both course credit (for the first 5 hours of participation, and $30 for the final 4 hours of participation). Previous research (e.g., Goska & Ackerman, 1996; Tomporowski, Simpson, & Hager, 1993) has indicated that such samples sometimes show differences on ability tests, whether through differences in motivation or in ability. With that as background, the hope was that the psychomotor tests

### Session 1

Paper & Pencil Ability Testing
1. Analogies (V)
2. Number Comparison (PS-Scanning)
3. Paper Folding (S)
4. Finding A/T (PS-Pattern)
5. Naming Symbols (PS-Memory)
6. Clerical Abilities 2 (PS-Scanning)
7. Vocabulary (V)
8. Coding (PS-Memory)
9. Verbal Test of Spatial (S)
10. Digit/Symbol (PS-Mem)
11. Problem Solving (N)
12. Number Comparison (PS-Scan)
13. Spatial Orientation (S)
14. Canceling Symbols (PS-Pattern)
15. Math Knowledge (N)

### Session 2

Paper & Pencil Ability Testing
1. Mechanical Knowledge (MK)
2. Controlled Associations (V)
3. Finding e/y (PS-Pattern)
4. Number Series (N)
5. Mechanical Comp (MK)

Computer (keyboard)
Choice/Simple RT Tests
1. 9-Choice RT
2. 4-Choice RT
3. 2-Choice RT
4. Simple RT

Computer (touchpanel)
Psychomotor Testing
[TouchPen input]
1. Single Tapping
2. Alternate Tapping
3. 8-Choice RT
4. 4-Choice RT
5. 2-Choice RT
6. Simple RT
7. 8-item Serial RT
8. 4-item Serial RT
9. Maze Tracing
10. Mirror Tracing

### Session 3

Computer (touchpanel)
Psychomotor Testing
[Finger input]
1. Single Tapping
2. Alternate Tapping
3. 8-Choice RT
4. 4-Choice RT
5. 2-Choice RT
6. Simple RT
7. 8-item Serial RT
8. 4-item Serial RT
9. Maze Tracing
10. Mirror Tracing

Figure 5. The presentation order of tests across the three sessions of Experiment 2. S = Spatial, V = Verbal, N = Numerical, Mk = Mechanical, PS = Perceptual Speed.

would be relatively robust in mean and variance levels, even with somewhat different participant populations. To evaluate this issue, means and standard deviations were computed for each test, and compared via independent *t*-tests (with the strategy that would most likely detect differences, that is, by using a per comparison $\alpha = .05$ -- which increases the experiment-wide Type I error rate). **Table 7** shows the respective means, sd's and *t*-test results for the two experiments. Of the 20 comparisons, five showed significant differences, favoring the Experiment 2 examinees. The table shows that, moreover, with the exception of the Single Tapping-Finger, the difference in means was also associated with substantially smaller between-subject variabilities in Experiment 2 -- suggesting that the Introductory Psychology student sample was also more homogeneous in ability than was the paid examinee sample. All in all, though, these differences, while significant with 232 degrees of freedom, tended to be modest, from a meaningfulness perspective. The largest difference between groups (the 8-item Serial RT test, was less than $.4\sigma$). Correlations between TouchPen and Finger administrations of the same tests were remarkably similar from Experiment 1 to Experiment 2, but are not shown here, for the sake of brevity.

Maze Tracing and Mirror Tracing. Given the increased complexity of these two tests (partly indicated by the substantially longer completion times), it comes as no surprise that significant performance improvements across the 20 trials in each condition were found. However, a full analysis of practice effects is not possible with these data, given that the stimulus maze (or mirror pattern) was changed every 5 trials (which had the effect of initially increasing completion time). Nonetheless, completion time on the TouchPen version of the Maze Tracing test improved from M = 19.81 sec, sd = 4.80 sec. (for the first block of 5 trials) to M = 13.61 sec, sd = 2.87 sec in the last block of 5 trials ($t(115) = 11.86$, $p < .01$). (Somewhat shallower learning effects were found for the Finger version, M = 18.35 sec., sd = 4.45 sec in the first five trials, and M = 15.9 sec, sd = 3.2 sec, $t(115) = 7.77$, $p < .01$.) In the TouchPen version of the Mirror Tracing test, initial block performance was M = 49.19 sec, sd = 20.73 sec, and final block performance was M = 33.13 sec, sd = 9.22 sec ($t(115) = 7.48$, $p < .01$). (Significant learning effects were found for the Finger version, M = 35.69 sec., sd = 9.30 sec in the first five trials, and M = 26.27 sec, sd = 5.62 sec, $t(114) = 15.85$, $p < .01$.) Scores used in the analyses below are averaged completion times across all four blocks of practice on each test.

Psychomotor Test Construct Validity. The array of ability measures administered in Experiment 2 provides an opportunity to examine the psychomotor tests in a broad ability nomothetic network. Tests of four major ability types were included (Verbal, Spatial, Numerical, and Mechanical), along with a series of finely graded perceptual speed measures (to assess three families of Perceptual Speed ability -- namely, Scanning, Pattern Recognition, and Memory). Two relatively coarse approaches were taken to place the psychomotor measures in the ability construct space -- multidimensional scaling and factor analysis. A third, more precise approach, compared correlations between the three PS composites and each of the psychomotor test variables. Each of these will be presented in turn below.

Table 7. Means, standard deviations, and t-tests for Experiment 1 and Experiment 2 psychomotor touchpanel tests.

| | $M_{\text{Experiment 1}}$ | $sd_{\text{Experiment 1}}$ | $M_{\text{Experiment 2}}$ | $sd_{\text{Experiment 2}}$ | $t$ |
|---|---|---|---|---|---|
| Keyboard 9Choice RT | 776 | 150 | 740 | 137 | 1.91* |
| Keyboard 4Choice RT | 545 | 77 | 534 | 67 | 1.16 |
| Keyboard 2Choice RT | 381 | 52 | 373 | 40 | 1.31 |
| Keyboard Simple RT | 255 | 52 | 251 | 42 | .64 |
| TouchPen 8Choice RT | 550 | 57 | 555 | 52 | -.70 |
| TouchPen 4Choice RT | 504 | 59 | 509 | 46 | -.72 |
| TouchPen 2Choice RT | 493 | 59 | 498 | 45 | -.72 |
| TouchPen Simple RT | 470 | 61 | 470 | 49 | 0.00 |
| Finger 8Choice RT | 583 | 66 | 580 | 60 | .36 |
| Finger 4Choice RT | 540 | 76 | 536 | 64 | .43 |
| Finger 2Choice RT | 527 | 81 | 521 | 63 | .63 |
| Finger Simple RT | 505 | 80 | 496 | 70 | .91 |
| TouchPen 8 Serial RT | 4094 | 564 | 3882 | 427 | 3.23** |
| TouchPen 4 Serial RT | 1667 | 204 | 1630 | 174 | 1.49 |
| Finger 8 Serial RT | 3897 | 612 | 3689 | 433 | 2.99** |
| Finger 4 Serial RT | 1649 | 218 | 1605 | 188 | 1.65* |
| TouchPen Single Tapping | 93 | 12 | 90 | 16 | 1.62 |
| Finger Single Tapping | 90 | 11 | 87 | 13 | 1.90* |
| TouchPen Alternate Tapping | 35 | 6 | 35 | 6 | 0.00 |
| Finger Alternate Tapping | 35 | 6 | 34 | 5 | 1.38 |

NOTE: Positive $t$ values indicate that Experiment 2 mean performance was "better."

Factor Analysis. The cognitive ability battery included previously standardized and validated tests to define four broad factors: Verbal, Math, Spatial, and Mechanical abilities. Three tests were administered for each of the four factors, except for the Mechanical factor, which was composed of only two tests. In addition, based on research with an extensive battery of Perceptual Speed (PS) measures (Ackerman & Rolfhus, 1996), three tests each were administered to assess a PS-Memory factor, a PS-Pattern Recognition factor, and a PS-Scanning factor. Together, there were 20 paper and pencil ability tests, administered over 4½ hours of testing. These tests were added to the 14 psychomotor tests (keyboard and TouchPen only, to avoid spurious factors that might have occurred if the Finger versions of the touchpanel tests were included, given the respective high communalities), for a total correlation matrix of 34 variables. The Humphreys-Montanelli (1975) parallel analysis criterion was used for the determination of the number of factors to extract. The parallel analysis indicated that 6 factors should be extracted. A principal axis factor analysis was used, with iterated communalities and squared multiple correlations as initial communality estimates. The solution was rotated to an oblique criterion by the Tucker-Finkbeiner (1981) Direct Artificial Personal Probability Factor Rotation (DAPPFR) algorithm (one-sided, given the positive manifold manifest in nearly all ability data). The DAPPFR-obtained solution is shown in **Table 8.**

The factor solution appears to be quite informative, in terms of evaluating how the various measures hang together in a battery of tests that is much broader in scope than is typically seen in the literature (though see one notable exception in Allison, 1960; the re-analyses of some of his data by Snow, Kyllonen, & Marshalek, 1984; and a more complete reanalysis by Ackerman, 1988). In this case, the cognitive ability tests tend to all hang together (e.g., all of the broad ability tests load saliently on Factor III, defined as "cognitive ability"), as would be expected, since these abilities have more variance in common with one another than they do with the perceptual speed and psychomotor factors. However, the Verbal ability tests also load significantly on Factor V (as do two of the perceptual speed tests that use substantial verbal content) -- this factor was defined as the "Verbal" factor (as distinct from the broader cognitive ability factor). The PS tests similarly all load highly on a single factor -- Factor II, which we identified as "PS." However, examination of the factor pattern matrix indicates that the three families of PS tests have respectively different configurations of loadings on other factors. The PS-Memory tests also load on Factor VI, which has significant loadings of two spatial tests that have high demands on working memory capacity (namely the Paper Folding test, which requires that the examinee keep track of a series of sequential paper folds; and the Verbal Test of Spatial Ability, which requires that the examinee visualize entire spatial problems presented orally, without resorting to note-taking -- see Ackerman & Kanfer, 1993 for an extensive discussion of these measures). The PS-Pattern Recognition tests appear to 'define' the broad PS factor, in that they have the highest loadings on that factor, and no significant loadings on other factors.

Turning to the psychomotor tests, Factor I appears to be well-defined by the entire set of touchpanel choice, simple and serial RT tests, whereas the keyboard Choice and Simple RT tests define their own factor (Factor IV). On the other hand, the TouchPen Single and

Table 8. Oblique Factor Solution (Pattern Loadings and Factor Intercorrelation Matrix) to Cognitive, Perceptual Speed, and Psychomotor Tests.

| | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| Verbal | | | | | | |
| Verbal Analogy | .025 | -.056 | .416 | .070 | **.409** | .252 |
| Controlled Associations | .010 | .230 | **.470** | -.044 | **.424** | -.047 |
| Extended Range Vocabulary | -.059 | .041 | **.499** | -.087 | **.667** | -.020 |
| Math | | | | | | |
| Math Knowledge | .031 | .099 | **.417** | .123 | .274 | .138 |
| Problem Solving | -.083 | -.095 | **.566** | .169 | .201 | .146 |
| Number Series | -.086 | .094 | **.455** | **.317** | .157 | .072 |
| Spatial | | | | | | |
| Paper Folding | -.023 | .026 | **.485** | .076 | .022 | **.366** |
| Verbal Test of Spatial Ability | .047 | -.032 | **.480** | .018 | -.111 | **.338** |
| Spatial Orientation | .045 | .216 | **.457** | -.036 | .061 | .238 |
| Mechanical | | | | | | |
| Mechanical Knowledge | .028 | .018 | **.561** | -.039 | .091 | .088 |
| Mechanical Reasoning | .028 | -.044 | **.761** | .004 | .058 | .249 |
| PS-Memory | | | | | | |
| Coding | -.060 | **.479** | -.020 | -.081 | .035 | **.451** |
| Naming Symbols | .059 | **.460** | .008 | .097 | .088 | **.458** |
| Digit/Symbol Substitution | -.058 | **.560** | .009 | .053 | -.029 | **.506** |
| PS-Pattern Recognition | | | | | | |
| Finding A/T | -.056 | **.770** | -.031 | -.028 | .069 | -.079 |
| Finding ∈/¥ | -.034 | **.748** | .034 | .125 | -.073 | -.079 |
| Canceling Symbols | .054 | **.607** | .220 | -.202 | -.056 | .136 |
| PS-Scanning | | | | | | |
| Name Comparison | .061 | **.665** | -.017 | .019 | **.436** | -.047 |
| Number Comparison | .088 | **.607** | -.109 | .027 | .233 | .003 |
| Clerical Abilities-2 | -.007 | **.554** | .077 | .110 | **.441** | .089 |
| Psychomotor Tests | | | | | | |
| TouchPen Single Tapping | .275 | .051 | .245 | .226 | -.140 | .026 |
| TouchPen Alternate Tapping | **.487** | .111 | .064 | .064 | -.267 | .105 |
| Keyboard 9-Choice RT | .050 | -.016 | -.031 | **.601** | .068 | .096 |
| Keyboard 4-Choice RT | .079 | -.017 | -.029 | **.714** | .050 | .094 |
| Keyboard 2-Choice RT | .213 | .033 | .029 | **.511** | .058 | -.018 |
| Keyboard Simple RT | .245 | .001 | .131 | **.382** | -.135 | -.119 |
| TouchPen 8-Choice RT | **.753** | -.108 | .022 | -.068 | .012 | -.039 |
| TouchPen 4-Choice RT | **.745** | -.046 | -.068 | -.011 | .102 | .047 |
| TouchPen 2-Choice RT | **.728** | .015 | -.110 | .055 | .052 | -.024 |
| TouchPen Simple RT | **.648** | .022 | .025 | .069 | .042 | -.113 |
| TouchPen 8-Serial RT | **.540** | .218 | .093 | -.067 | .025 | .097 |
| TouchPen 4-Choice RT | **.587** | .099 | .034 | -.006 | .062 | .084 |
| TouchPen Maze Tracing | .195 | -.055 | **.312** | .084 | -.041 | -.194 |
| TouchPen Mirror Tracing | .070 | -.022 | **.487** | .242 | -.036 | -.195 |

| Primary Factor Correlations | I | II | III | IV | V |
|---|---|---|---|---|---|
| Factor II | **.305** | | | | |
| Factor III | .248 | .020 | | | |
| Factor IV | **.464** | .177 | .105 | | |
| Factor V | .022 | .009 | -.214 | .087 | |
| Factor VI | .010 | .051 | .012 | .205 | .277 |

NOTE: salient loadings in **bold**.

Alternate Tapping tests fit less well in the factor solution (possibly because so few measures of this ability were included) -- but Alternate Tapping loads highly with the other discrete psychomotor tests (Factor I), and single tapping also has its highest loading on this factor, though not significantly. Finally, the Maze Tracing and Mirror Tracing tests clearly load significantly only on Factor III, the broad Cognitive Ability factor, and not on Factor I (the Discrete Psychomotor Ability factor). Mirror tracing, in fact, has as high a loading on the Cognitive Ability factor as many of the tests that were specifically designed to tap high-level cognitive processes. From this analysis, it appears that these two continuous-response psychomotor tests have more in common with the paper & pencil tests than they do with the discrete-response psychomotor tests, something that was not readily apparent from examining only the matrix of raw intercorrelations.

Finally, significant correlations were obtained between Factor I and Factor II, and between Factor I and Factor IV, which similarly fits well with the conceptualization of the nature of these ability constructs. That is, Discrete Response Psychomotor abilities (Factor I) share common variance with both a broad PS ability (Factor II), and with a Discrete Psychomotor ability captured via keyboard assessment of Choice and Simple RT (Factor IV).

Multidimensional Scaling. Factor analysis is a method that attempts to extract estimates of hypothetical constructs underlying a set of observed tests. Another way to conceptualize the relations among the various constructs is to consider that correlations can serve as proxies for similarity estimates (tests that are highly correlated with one another will be close together in construct space, and tests that have a zero correlation will be far away from one another). Although computationally similar to factor analysis, multidimensional scaling (MDS) can be used to provide a more visually tractable representation of the relative proximities among a set of test measures (see, for example the discussion and demonstration by Marshalek, Lohman, and Snow, 1983). Traditionally, when one applies a radex approach (Guttman, 1954; Snow, et al., 1984) to an ability battery, the resulting MDS solution (in two or three dimensions) can provide a means toward evaluating how similar (or different) various measures are from one another. In is important to note, though, that the MDS approach is mathematically equivalent to a hierarchical factor analysis procedure (see Marshalek, et al., 1983 for a discussion and illustrations; see also Ackerman, 1988; Ackerman et al., 1995).

The starting point, then, for this analysis was the same correlation matrix used in the factor analysis discussed above. Given that a few of the correlations were negative (even though none significantly), a constant (1) was added to all correlations, so that all similarly values were positive. The matrix was then subjected to KYST-3 multidimensional scaling (MDS) (Kruskal, Young, & Seery, 1973), a two-dimensional solution was extracted (Stress Formula 1 = .174), and as is customary, the solution was rotated to a principal-components orientation. The solution is plotted in **Figure 6**.

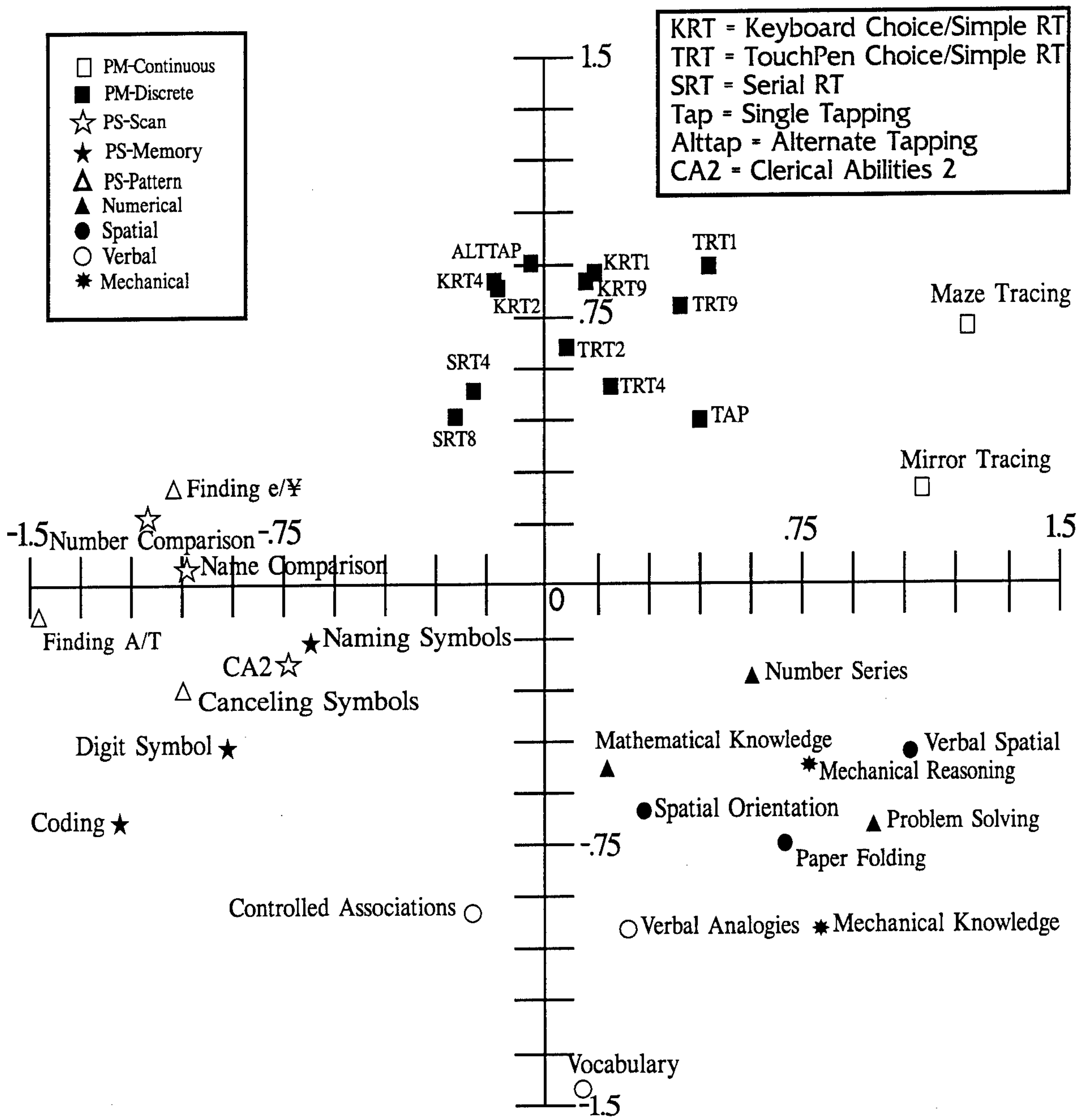


Figure 6. Multidimensional scaling (KYST3) solution to the tests administered in Experiment 2. Different symbols represent different a priori ability factors.

The MDS solution provides an important illustration of several aspects of the data. First, each family of tests (Cognitive, PS, and Psychomotor) occupies a well-defined separate region of the MDS space. Also, tests that share content (such as the verbal ability tests) tend to be closer to one another than to other tests, but with a few exceptions. One exception is that the mechanical ability tests tend to be closer in proximity to some spatial tests than to one another. Also, the math problem solving test was more closely related to the spatial tests, which also have high working memory demands than to the other math tests. The PS tests, while defining a unique region of the solution space, nonetheless are sometimes distinct (such as the Coding test), and sometimes close in proximity. Nonetheless, despite the Keyboard and touchpanel Choice/Simple RT tests defining separate factors in the factor analysis, the significant correlation between those respective factors is concordant with the result that all of these tests are in close proximity with one another. Conversely, the Serial RT tests and the Single Tapping test are a little removed from the other discrete response psychomotor tests. Finally, the Maze Tracing and Mirror Tracing tests are clearly distinct from the other psychomotor tests, and from one another. The Mirror Tracing test has nearly closer proximity to the cognitive ability tests than it does to the other psychomotor tests.

PS and Psychomotor Tests. To assess whether the more specific PS abilities showed differential relations with the psychomotor abilities, three PS ability composites were formed (the composites are sums of unit-weighted *z*-scores for each of the three tests for each PS ability). Correlations were then computed between the PS ability composites and the psychomotor tests. Briefly, it was found that the Choice and Simple RT tests tended to be most highly associated with the PS-Scanning ability (mean $r$ = .30), but less associated with PS-Pattern recognition (mean $r$ = .18) and PS-Memory (mean $r$ = .21). In contrast, the Serial RT tests correlated significantly with all three PS composites, most highly with PS-Scanning (mean $r$ = .43, .30, and .31 for PS-Scanning, PS-Pattern, and PS-Memory, respectively), which is consistent with our a priori notions of the strategies that are used by the examinees in making the multiple responses associated with these tests. The tapping and alternate tapping tests showed small, but undistinguished correlations with all three PS factors (mean $r$ = .22, .24, and .25). Finally, the Maze and Mirror Tracing tests failed to show any significant relations with the three PS abilities (mean $r$ = .07, .08, and .03) -- again suggesting that these tests capture variance more in common with broad cognitive abilities than with either discrete response psychomotor abilities or with perceptual speed abilities.

Experiment 2 Discussion

The touchpanel tests developed and explored in Experiment 1 yielded results that were highly consistent in Experiment 2 -- indicating the general robustness of the methodology and the test designs. The Maze Tracing and Mirror Tracing tests, while presenting two additional technological challenges (measurement of continuous responses in both tests, and providing a means toward assessing 'mirrored' response dynamics in the Mirror Tracing test), yielded a potentially critical link between simple psychomotor tests on the one hand (the discrete response tests) and broader cognitively-related psychomotor abilities on the other hand. That linkage is possibly inherent in the fact that, especially for the Mirror

Tracing test, the examinee does not initially "know" what response to make in order to get the cursor to move in the intended direction. Thus, in contrast to the more discrete tests described here, the mirror tracing (and to a smaller degree, the maze tracing) test involves both psychomotor components and a non-trivial contribution of cognitive components.

The two multivariate analyses provided support for theories of ability that include considerations of psychomotor ability, along with cognitive and perceptual speed abilities (e.g., Ackerman, 1988). The more substantial demands on cognitive processes placed in the maze and mirror tracing tests, while providing a linkage between discrete-response psychomotor abilities and cognitive abilities, also show substantial variance that is not common to either set of ability factors, suggesting that such abilities can be effectively separately identified and assessed. The fact that the discrete-response psychomotor tests defined a factor that shared little variance with the broad cognitive ability factor provides an additional demonstration that such tests are capable of adding incremental validity to prediction of real-world performance measures -- though the demonstration of such validity (in the context of broad perceptual speed ability predictors) was partly demonstrated in Experiment 1. To a lesser extent, the Serial RT tests provide a more direct linkage to the Perceptual Speed ability domain than is found with the Choice and Simple RT tests, especially with the PS-Scanning domain. All of the discrete-response psychomotor tests, though, have much common variance -- such that it may be possible to obtain useful assessment of this psychomotor ability with a relatively brief battery of tests, especially because reliable and valid psychomotor ability assessment can be performed with much shorter tests than are typically used in the broader cognitive domain.

## V. Experiment 3-A

The main goal for Experiment 3-A was to try out and evaluate the final two tasks in the suite of touch-panel psychomotor tests -- tasks that involved continuous tracking performance, namely the Maze Pursuit task and the Rotary Pursuit task.

<u>Development of New Touch-Panel Psychomotor Tests</u>. Two new touchpanel tests were added to the extant battery: Maze Pursuit and Rotary Pursuit.

<u>Maze Pursuit</u>. The Maze Pursuit Test is a relatively straightforward theoretical extension of the Maze Tracing test used in the prior two studies. The main goal of measuring a examinee's ability to follow a moving target through a maze turned out to be a rather substantial programming challenge. The first challenge was to make the target move in a fluid motion through the maze. This was accomplished by redrawing the target quickly, displaced by a few pixels at a time. The second challenge was to provide a target that could be seen by the examinee while the examinee simultaneously had the touch-pen placed over the target. This was accomplished by using a circular target and a slightly wider (roughly 20% larger) track for the maze. The last challenge was to derive an appropriate measure of task performance. This was accomplished by drawing on the apparatus version of pursuit tasks -- established as time-on-target, rather than a root-mean-square error, which could have

erroneously identified as "good" performance if a examinee had never had the touch-pen "on target," but had closely trailed the target throughout the trial. After extensive pilot testing, the speed of the target through the maze was set so that it would travel from start to finish in about 17 seconds, depending on the specific maze pattern. Time-on-target was the number of seconds (out of the total traversal time) that the touch-pen was placed on the screen in the same place/time as the target.

Rotary Pursuit. The Rotary Pursuit test followed the Maze Pursuit Test in an analogous fashion, including the circular target, the "home key," and the time-on-target performance criterion -- see **Figure 7**. The only substantive difference was that the Rotary Pursuit test had a constant circular track, instead of a maze. Each Rotary Pursuit trial lasted 20 seconds.

Method

Participants. One hundred thirty-one adults ($\underline{M}_{age}$ = 21.52, sd = 3.26 years) participated in Experiment 3 -- 88 women and 43 men. The same inclusion criteria were used as in the previous experiments (namely, age between 18 and 30 years, normal or corrected-to-normal vision, hearing, and motor coordination).

With the exception of information provided below, all other details of the study (e.g., apparatus, procedural matters, and so on) were identical to those of Experiments 1 and 2.

Ability Tests. Pencil-and-paper testing included tests to assess the following ability factors:

1. Mathematical: Problem Solving, Mathematical Knowledge, Number Series
2. Spatial: Paper Folding, Verbal Test of Spatial Ability, Spatial Orientation
3. Perceptual Speed-Complex: Dial Reading, Directional Headings
4. Perceptual Speed-Scanning: Number Comparison, Clerical Abilities-2, Name Comparison
5. Perceptual Speed-Pattern Recognition: Finding A/T, Canceling Symbols, Finding ∈/¥
6. Perceptual Speed-Memory: Naming Symbols, Coding, and Digit Symbol

Procedure. Experiment 3-A was conducted in two sessions that are illustrated in **Figure 8**. In Session 1, the first set of paper and pencil ability testing was followed by the entire suite of touch-panel psychomotor tests with touch-panel input. The touch-panel tests were identical to those in Experiment 2, except for the addition of 3 blocks of 10 trials of the Maze Pursuit test, and 20 trials of the Rotary Pursuit test. In Session 2, the remaining paper and pencil ability tests were administered, followed by re-testing of the Maze Pursuit test (3 blocks of 10 trials) and the Rotary Pursuit test (20 trials).

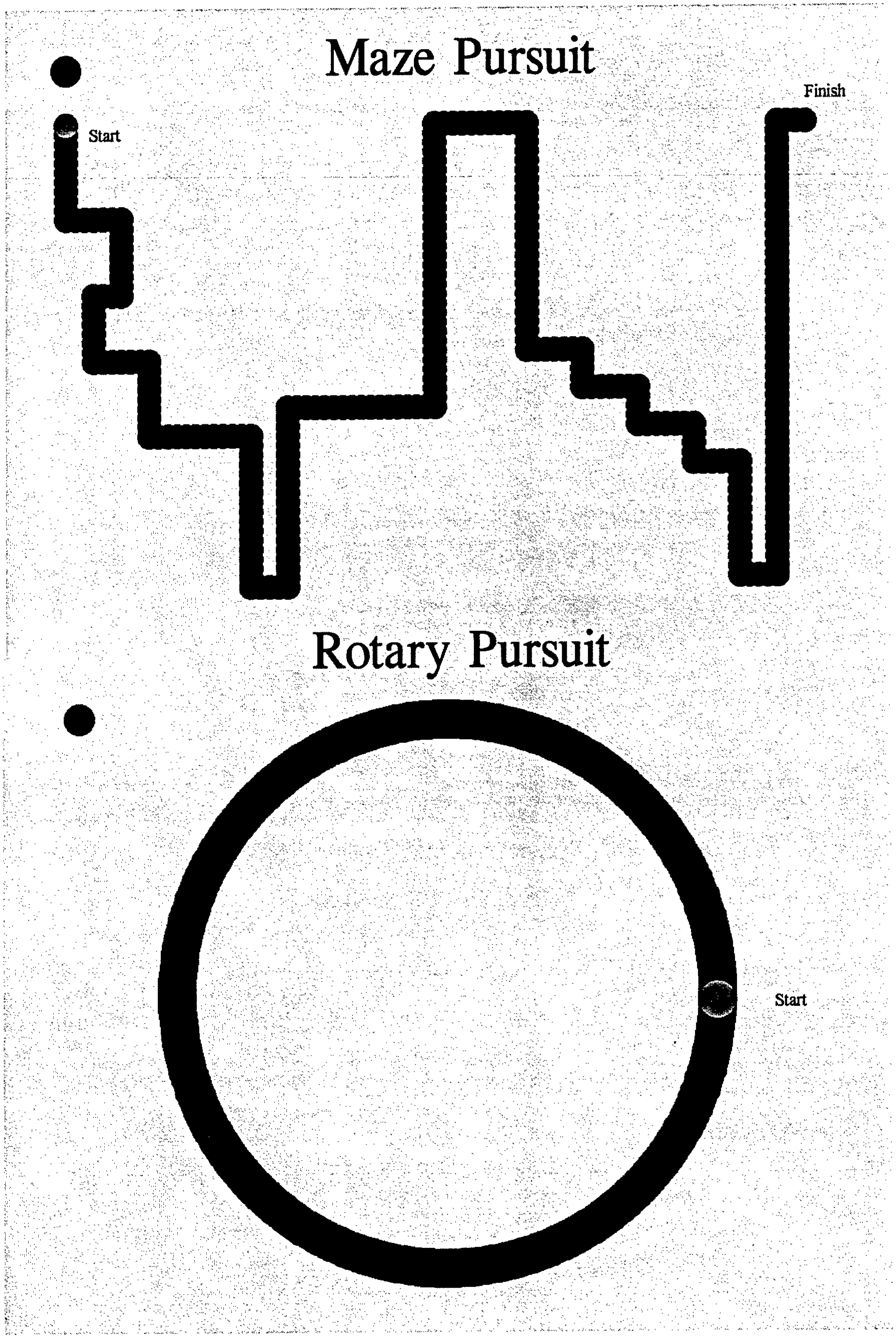


Figure 7. Bitmapped displays for the Maze Pursuit (top panel) and the Rotary Pursuit (bottom panel) for Experiment 3-A.

### Experiment 3-A
### Session 1

Paper & Pencil Ability Testing
1. Problem Solving (N)
2. Number Comparison (PS-Scanning)
3. Paper Folding (S)
4. Finding A/T (PS-Pattern)
6. Clerical Abilities 2 (PS-Scanning)
7. Dial Reading (PS-Complex)

Computer (touchpanel)
Psychomotor Testing
1. Single Tapping
2. Alternate Tapping
3. 8 Choice RT
4. 4 Choice RT
5. 2 Choice RT
6. Simple RT
7. 8- item Serial RT
8. 4-item Serial RT
9. Maze Tracing
10. Mirror Tracing
11. Maze Pursuit
12. Rotary Pursuit

### Experiment 3-A
### Session 2

Paper & Pencil Ability Testing
1. Digit/Symbol (PS-Memory)
2. Verbal Test of Spatial (S)
3. Canceling Symbols (PS-Pattern)
4. Math Knowledge (N)
5. Naming Symbols (PS-Mem)
6. CAB Mechanical Knowledge (MK)
7. Coding (PS-Memory)
8. Spatial Orientation (S)
9. Finding e/y (PS-Pattern)
10. Directional Headings (PS-Complex)
11. Number Series (N)
12. DAT Mechanical Knowledge (MK)

Computer (touchpanel)
Psychomotor Testing
1. Maze Tracing - retest
2. Mirror Tracing - retest
3. Maze Pusuit - retest
4. Rotary Pursuit - retest

### Experiment 3-B
### Session 3

Apparatus Tests
1. Maze Tracing
2. Rotary Pursuit
3. Mirror Tracing
4. Maze Tracing - retest
5. Rotary Pursuit - retest
6. Mirror Testing - retest

Barcode Scanning Work Sample task
1. Video instructions
2. Barcode scanning

Computer (touchpanel)
1. Single Tapping
2. Alternate Tapping

Barcode Scanning Work Sample task
1. Barcode scanning - retest

Figure 8. The presentation order of tests across the three sessions of Experiment 3-A and 3-B. S = Spatial, V = Verbal, N = Numerical, Mk = Mechanical, PS = Perceptual Speed.

Results

For the sake of brevity, results regarding the basic properties of the paper and pencil tests and the psychomotor tests administered in the previous experiments will be eliminated from the current discussion. Instead, the main focus will be on the reliability and validity of the two new pursuit tests.

Maze Pursuit. Performance on the Maze Pursuit test was found to be generally reliable, though there was an effect of practice from Session 1 to Session 2. Performance at Session 1 was M = 7.86 sec, sd = 1.84 sec and performance at Session 2 was M = 9.81 sec, sd = 1.79 sec. A paired *t*-test for the difference between means was significant ($t$ (130) = -30.97, $p < .01$), indicating that performance significantly improved from Session 1 to Session 2. However, the ordering of individual differences was consistent from Session 1 to Session 2, in that the two scores correlated $r = .92$, $p < .01$.

Rotary Pursuit. Similar to the Maze Pursuit test, the Rotary pursuit test showed both significant performance improvement from Session 1 to Session 2 (Session 1 M = 8.08 sec, sd = 2.11 sec; Session 2 M = 10.69 sec, sd = 2.00 sec.; $t$ (130) = -30.46, $p < .01$), and high test-retest reliability ($r = .89$, $p < .01$). Moreover, composite scores for the two pursuit tasks (across sessions) correlated very highly ($r = .84$), indicating that both tests were pretty much measuring the same underlying trait.

Validity of Pursuit Tests. For the assessment of construct validity, the two pursuit tests were correlated with the various paper and pencil ability composites, and with composites from the other touch-panel psychomotor tests. Results of these comparisons are shown in **Table 9**. In general, both pursuit tests showed substantial communality with cognitive, perceptual speed, and other psychomotor measures -- though none of the correlations was so high to indicate that the tests were substitutable. The pursuit tests were most highly correlated with Maze Tracing, Mirror Tracing, and Serial RT tests from the psychomotor domain, and Spatial ability and Mechanical Knowledge from the cognitive domain. The pursuit tests were less highly related to the three Perceptual Speed factors of Scanning, Pattern, and Memory. Also, the Maze Pursuit test tended to show slightly higher correlations with the various other measures than the Rotary Pursuit test.

## VI. Experiment 3-B

As part of another study (Field, 1998), participants from Experiment 3-A were invited back to the laboratory to complete a series of apparatus tests -- see **Figure 8** for the experiment layout. Seventy-nine examinees returned for a third and final session. During this session, they completed three sets of apparatus psychomotor tests, namely Apparatus Maze Tracing, Apparatus Mirror Tracing, and Apparatus Rotary Pursuit. In addition, they completed a synthetic work-sample task, called "Barcode Scanning." The Apparatus Maze Tracing test was the same apparatus used in Experiment 1. The other measures are described below.

Table 9. Experiment 3-A. Correlations between Ability Composites and Maze Pursuit, Rotary Pursuit Tests.

| | Maze Pursuit | Rotary Pursuit |
|---|---|---|
| Cognitive | | |
| Spatial Ability | .582** | .447** |
| Numerical Ability | .432** | .342** |
| Mechanical Knowledge | .534** | .444** |
| Perceptual Speed - Complex | .461** | .375** |
| Perceptual Speed | | |
| Perceptual Speed - Scanning | .247** | .228** |
| Perceptual Speed - Pattern | .272** | .322** |
| Perceptual Speed - Memory | .428** | .374** |
| Psychomotor | | |
| Single Tapping | .406** | .356** |
| Alternate Tapping | .508** | .444** |
| Choice RT | .425** | .408** |
| Serial RT | .581** | .537** |
| Maze Tracing | .579** | .489** |
| Mirror Tracing | .610** | .502** |

Note: ** $p < .01$

Apparatus Mirror Tracing. The apparatus was made by Marietta Apparatus Co., (Model #5-5). The task is to trace inside a star-shaped track with a stylus while monitoring progress reflected in a mirror placed perpendicular to the star. Monitoring of actual (as opposed to reflected) progress is prevented by a raised metal plate under which the star is placed. The width of the star is 13cm (5-3/16in). The depth (distance away from the viewer), is 15cm (6in). The width of the star track is 9mm (3/8in). Errors are recorded when the stylus comes into contact with the sides of the track, and total time-to-completion was recorded with a stopwatch.

Apparatus Rotary Pursuit. This apparatus was made by Lafayette Instrument Co. (Model #30010). The task is to hold a metal stylus to a circular target on a rotating turntable at a rate of 45 revolutions per minute. The diameter of the turntable is 25.5cm (10in), and the diameter of the target is 1.9cm (3/4in). The target is located near the edge of the turntable, and is identified by its silver color, as compared to the black color of the turntable. The stylus is spring-loaded to prevent the examinee from pressing down hard on the turntable and slowing the stimulus. Trials are 20 seconds, with a random start delay. Time-on-target is recorded as the time that the stylus is in physical contact with the circular target on the turntable.

Barcode Scanning Synthetic Work-Sample Task. The Barcode Scanning Task is essentially a simulation of the kinds of perceptual and psychomotor tasks that are used in some retail sales occupations. The task consisted of a market-basket full of 36 retail items (e.g., as a box of diskettes, a book, a box of binder clips, a card file box, etc.) and a Hewlett-Packard wand-style barcode reader, connected to a 80486 Compaq computer. The examinee was instructed to quickly and accurately pick up items from the basket, scan the barcode using the wand reader, and put the items into another basket. After a video instruction and demonstration, examinees were given 10 trials of the Barcode Scanning task. Performance was measured as the total time for scanning the 36 items and the number of scanning errors made (including duplicate scans and incomplete scans). Because total scanning time and errors were substantially correlated, $r = .59$, $p < .01$, a single unit-weighted $z$ score composite was created as an indicator of overall performance.

Results

The main results of this part of the study concerned the cross-correlations between the test battery from Experiment 3-A and the Apparatus Tests from Experiment 3-B. The correlations are shown in **Table 10**. The most salient findings from this study were that, for the most part, the highest correlations were found from the analog touch-panel and apparatus tests, namely: Maze tracing ($r = .50$, $p < .01$), Mirror Tracing ($r = .44$, $p < .01$), and Rotary Pursuit ($r = .50$, $p < .01$). While these correlations were not as high as the respective reliabilities would allow, the sense of these values is that the touch-panel tests were indeed substantially related to the apparatus tests.

Performance on the Barcode Scanning Task was not highly related to either the paper-and-pencil based ability tests or the touch-panel tests, but performance on the task was significantly related to key Perceptual Speed abilities and to the Maze Tracing composite.

Table 10. Experiment 3-B. Correlations between Touch-Panel Test Composites and Apparatus Tests.

| | Maze Tracing | Mirror Tracing | Rotary Pursuit | Barcode Scanning |
|---|---|---|---|---|
| Spatial Ability | .118 | .261* | .321** | .208 |
| Numerical Ability | .073 | .263* | .238* | .240* |
| Mechanical Knowledge | .156 | .387** | .254* | .210 |
| Perceptual Speed-Complex | .055 | .209 | .319** | .336** |
| Perceptual Speed- Scanning | .146 | .110 | .228* | .326** |
| Perceptual Speed-Pattern | .188 | .040 | .070 | .325** |
| Perceptual Speed-Memory | .210 | .162 | .349** | .138 |
| Single Tapping | .301** | .222 | .307** | -.028 |
| Alternate Tapping | .408** | .378** | .372** | .024 |
| Choice Reaction Time | .094 | .278* | .367** | .055 |
| Serial Reaction Time | .225 | .062 | .318** | .117 |
| Maze Tracing | **.500**** | .442** | .394** | .315** |
| Mirror Tracing | .162 | **.439**** | .312** | .099 |
| Maze Pursuit | .327** | .214 | .430** | .165 |
| Rotary Pursuit | .329** | .256* | **.498**** | .175 |

Note: Analog Test cross-correlations are shown in Boldface.
$*p < .05$; $**p < .01$.

A stepwise regression equation indicated that among all of the tests, Maze Tracing and Perceptual Speed - Patterns were most highly and significantly related to the Barcode Scanning Task, accounting for an initial 15% and an additional 10% of the total variance, respectively. Thus, 25% of the variance in the Barcode Scanning Task was accounted for by the touch-panel psychomotor test and the Perceptual Speed ability composite.

Experiment 3-A and 3-B Discussion

The main purpose of Experiment 3 was to assess the two final touch-panel psychomotor tests in the suite -- the Maze Pursuit and Rotary Pursuit tests. The results reported in Section 3-A indicate that both new tests were reliable and valid -- to the degree that these tests represent separable variance with respect to cognitive, perceptual speed, and the other psychomotor abilities. The two pursuit tests were found to be highly intercorrelated, suggesting that for most intents and purposes, they are substitutable measures of the same underlying ability. Section 3-B provided additional convergent validity evidence between the touch-panel and apparatus based tests, and also provided support for the external validity of at least one set of touch-panel tests (Maze Tracing) in predicting performance on a work-sample task.

## VII. Experiment 4

Overview

Experiment 4 was designed mainly to address the predictive validity of the full set of seven different touch-panel psychomotor tests, both in isolation and in conjunction with traditional paper and pencil measures of cognitive and perceptual abilities. Two criterion tasks were used for validation purposes, a predominantly speeded, consistent, skill acquisition task (the Kanfer-Ackerman ATC Task(c)) -- which was given extensive practice, and a complex, real-time problem solving and decision making task (TRACON). The Kanfer-Ackerman ATC task has been described in the context of Experiment 1 and Appendix A, and TRACON is described in Appendix B.

### Method

Participants

Ninety-eight adults between the ages of 18 and 30 years of age participated in the study. All examinees were native English speakers, and had normal or corrected-to-normal hearing, vision, and motor coordination. The final sample was made up of 53 women and 45 men, mean age = 20.8, sd age = 2.32, range 18-28 years.

Apparatus. Pencil-and-paper testing during Session 1-4 was administered in a laboratory with prerecorded instructions and directions presented over a public address system. Up to 12 examinees were tested at a time. Computerized testing during Sessions 2-6 was administered on Dell and IBM Pentium computers with 15" Microtouch touch-panel monitors at individual carrels.

Psychomotor Tests. The suite of touch-panel psychomotor tests includes all seven of the types of measures described in the previous experiments.

Ability Tests. Pencil-and-paper testing included tests to assess the following ability factors:

1. Verbal: MAB Comprehension, Extended Range Vocabulary, Controlled Associations, General Information
2. Spatial: Paper Folding, Verbal Test of Spatial Ability, Spatial Orientation
3. Mathematical: Problem Solving, Mathematical Knowledge, Number Series
4. Perceptual Speed-Complex: Dial Reading, Directional Headings
5. Perceptual Speed-Scanning: Number Comparison, Clerical Abilities-2, Name Comparison
6. Perceptual Speed-Pattern Recognition: Finding A/T, Canceling Symbols, Finding ∈/¥
7. Perceptual Speed-Memory: Naming Symbols, Coding, and Digit Symbol

TRACON®. The first criterion task was a real-time problem solving simulation of many of the tasks performed by air traffic controllers -- called TRACON for Terminal Radar Approach Controller. Details of the task are provided in Appendix B. This criterion was expected to demonstrate both convergent and discriminant validity for the various psychomotor ability measures. Simple psychomotor tests (such as Tapping and Choice RT) were expected to have modest correlations with TRACON performance, while the more complex, spatially involved psychomotor measures (e.g., Mirror Tracing) were expected to have more substantial correlations with TRACON performance. Ten 30-minute trials (five hours) of practice were given for the TRACON task -- substantially less than would be required for asymptotic skilled performance to be acquired, but sufficient practice to allow for stable ability-performance correlations to be revealed.

Kanfer-Ackerman ATC Task©. The ATC task (also used in Experiment 1) was used as the second criterion task. However, in this instantiation of the task, 32 ten-minute trials of practice (5 hours, 20 minutes) were provided with the expectation that asymptotic, mostly automatized performance levels would be achieved by most examinees. In accordance Ackerman's (1988) theory of the ability determinants of skilled performance, the new touch-panel psychomotor ability measures were expected to show substantial correlations with performance on the ATC task, especially when high levels of performance skill have been attained.

Procedure

The study took place over six sessions, totaling 24 hours. (See **Figure 9** for the layout for Experiment 4.) Session 1 consisted entirely of pencil-and-paper test administration and computerized psychomotor testing. Session 2 consisted of one hour of pencil-and-paper testing; one hour of computerized psychomotor testing; and a one-hour TRACON instructional video. For Session 3, .5 hour of ability testing was followed by 5, 30-minute TRACON trials. Session 4 was similar to Session 3, including .5 hour of ability testing and an additional 5, 30-minute TRACON trials. Session 5 and Session 6 were devoted to the Kanfer-Ackerman ATC task; instructions were given in Session 5, and each of these two sessions included 16, 10-minute task trials. Thus, a total of 5 hours time-on-task was given in TRACON, and 5 hours, 20 minutes time-on-task for the ATC task.

**Session 1**

Paper & Pencil Ability Testing

1 Problem Solving (N)
2. Number Comparison (PS-Scanning)
3. Paper Folding (S)
4. Finding A/T (PS-Pattern)
5. Clerical Abilities 2 (PS-Scanning)
6. Dial Reading (PS-Complex)

Computer (touchpanel) Psychomotor Testing

1. Single Tapping
2. Alternate Tapping
3. 8 Choice RT
4. 4 Choice RT
5. 2 Choice RT
6. Simple RT
7. 8- item Serial RT
8. 4-item Serial RT
9. Maze Tracing
10. Mirror Tracing
11. Maze Pursuit
12. Rotary Pursuit

**Session 2**

Paper & Pencil Ability Testing

1. Vocabulary (V)
2. Digit/Symbol (PS-Mem)
3. Number Series (N)
4. Canceling Symbols (PS-Scan)
5. General Information (V)
6. Verbal Test of Spatial (S)
7. Name Comparison (PS-Scan)
8. Math Knowledge (N)

Computer (touchpanel) Psychomotor Testing

1. Maze Tracing (retest)
2. Mirror Tracing (retest)

TRACON - Training Video

**Session 3**

Paper & Pencil Ability Testing

1. MAB Comrehension (V)
2. Find e/y (PS-Scan)
3. Controlled Associations (V)

TRACON — Five, 30 min. trials

**Session 4**

Paper & Pencil Ability Testing

1. Coding (PS-Mem)
2. Spatial Orientation (S)
3. Naming Symbols (PS-Mem)
4. Directional Headings (PS-Complex)

TRACON — Five, 30 min. trials

**Session 5**

Kanfer-Ackerman ATC Task Instructions

ATC Task — Sixteen 10-minute trials

**Session 6**

ATC Task — Sixteen 10-minute trials

Figure 9. The presentation order of tests and criterion tasks across the six sessions of Experiment 4. S = Spatial, V = Verbal, N = Numerical, PS = Perceptual Speed.

Results

While there are many different ways of assessing the data from Experiment 4, the current report will focus on the unique aspects of the study in comparison to the previous three experiments. In particular, the results will be divided into five sections, as follows: (1) A brief review of the learning/performance data from the TRACON and ATC criterion tasks over task practice; (2) A review of correlates between a general ability composite and performance on the two criterion tasks; (3) A review of the relations among predictor measures, including the cognitive, perceptual speed, and psychomotor ability domains; (4) A review of cross-correlations between predictor composites and initial and final performance on the criterion tasks, and (5) A multiple correlation evaluation of the incremental predictive validity of the new touch-panel psychomotor tests.

TRACON and ATC Performance over Practice. **Figure 10** shows the mean and between-subject *sd* measures across task practice on the two criterion tasks. For TRACON, each 30-minute trial is indicated as a separate data point. In contrast, for the ATC task, each point represents the mean performance for four 10-minute trials (i.e., 40-minutes of time-on-task). As can be easily seen from both figures, practice resulted in a substantial increase in task performance. Mean performance on TRACON at Trial 1 was 6.92 planes handled, and for Trial 10, mean performance was 14.55 planes handled, a highly significant increase in performance ($t$ (94) -12.08, $p < .001$). Mean performance on the ATC task Session 1 was 43.63 planes landed, and for Session 8, mean performance was 66.12 planes landed, also a highly significant increase in performance ($t$ (97) -24.86, $p < .001$). In contrast, the between-subjects *sd* measures indicate a different pattern of the effects of practice. For TRACON Trial 1, $sd = 4.67$, and for Trial 10, $sd = 6.97$, an increase in *sd* of 49% -- indicating that spread of individual differences in performance increases substantially with practice. In contrast, for ATC Session 1, $sd = 11.33$ and for Session 8, $sd = 9.33$, a decrease of 17% -- indicating that individuals become more alike after practice, and there is thus less variance in performance available to be explained by predictor measures.

Correlations between General Ability and Criterion Task Performance. **Figure 11** shows the correlations between a traditional composite of general intellectual ability (Numerical + Verbal + Spatial abilities) and the two criterion tasks. As can be clearly seen from the figure, TRACON represents stable demands on general ability over the five hours of task practice, while the ATC task shows substantially diminishing demands on general ability with task practice. These two findings match very closely with Ackerman's theory of the ability determinants of performance during skill acquisition (e.g., Ackerman, 1986, 1987, 1988) and earlier empirical investigations (e.g., Ackerman, 1990, 1992; Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995) with these and other tasks. That is, the TRACON task represents a class of tasks with continued novel information processing, problem solving and decision making demands, while the ATC task represents a class of tasks with initially demanding information processing components, but with consistent characteristics that allow for the development of automatic processes over practice, which in turn, result in diminished demands on general intellectual abilities.

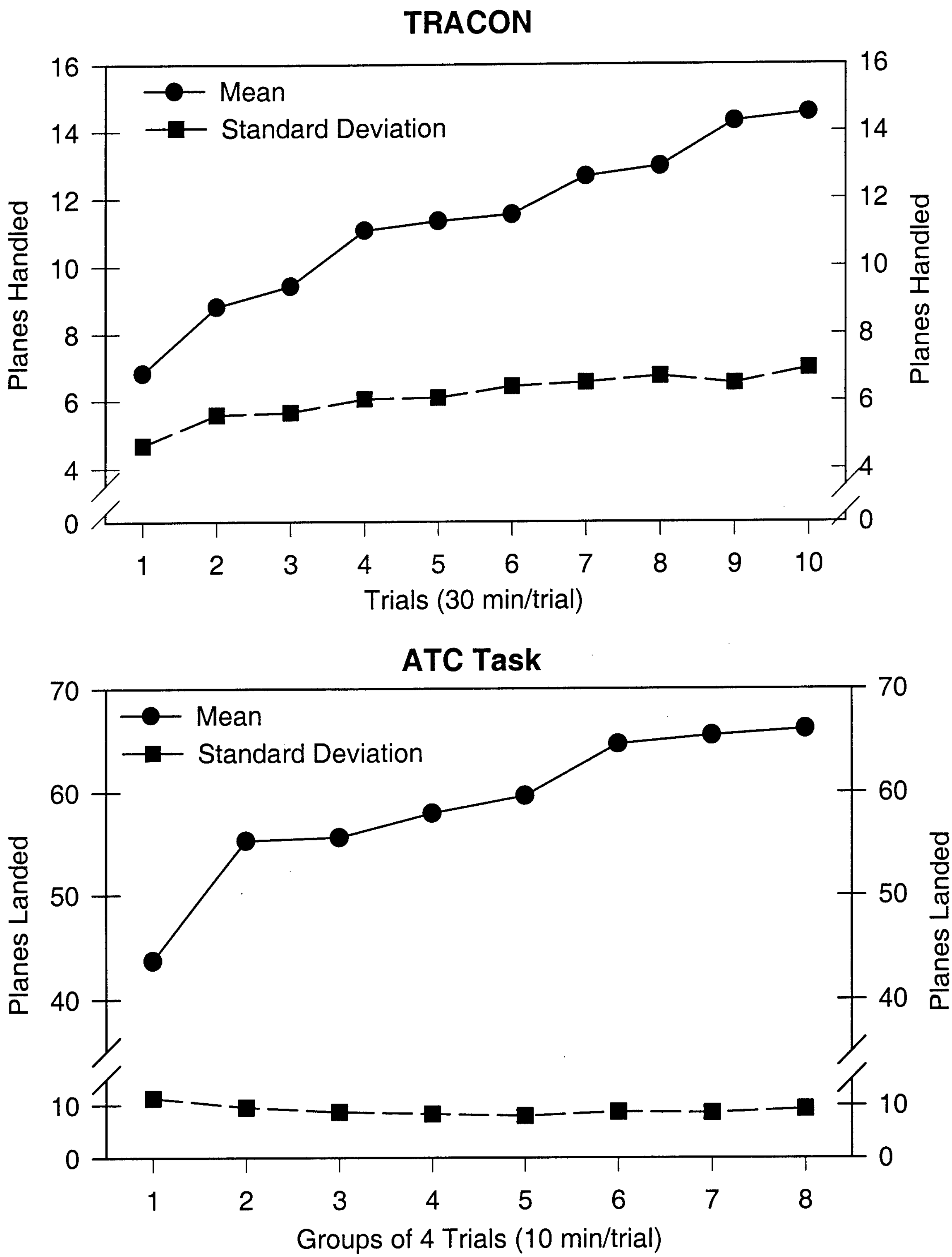


Figure 10. Mean performance and between-subject standard deviations as a function of task practice for TRACON(r) and the Kanfer-Ackerman ATC (c) Task.

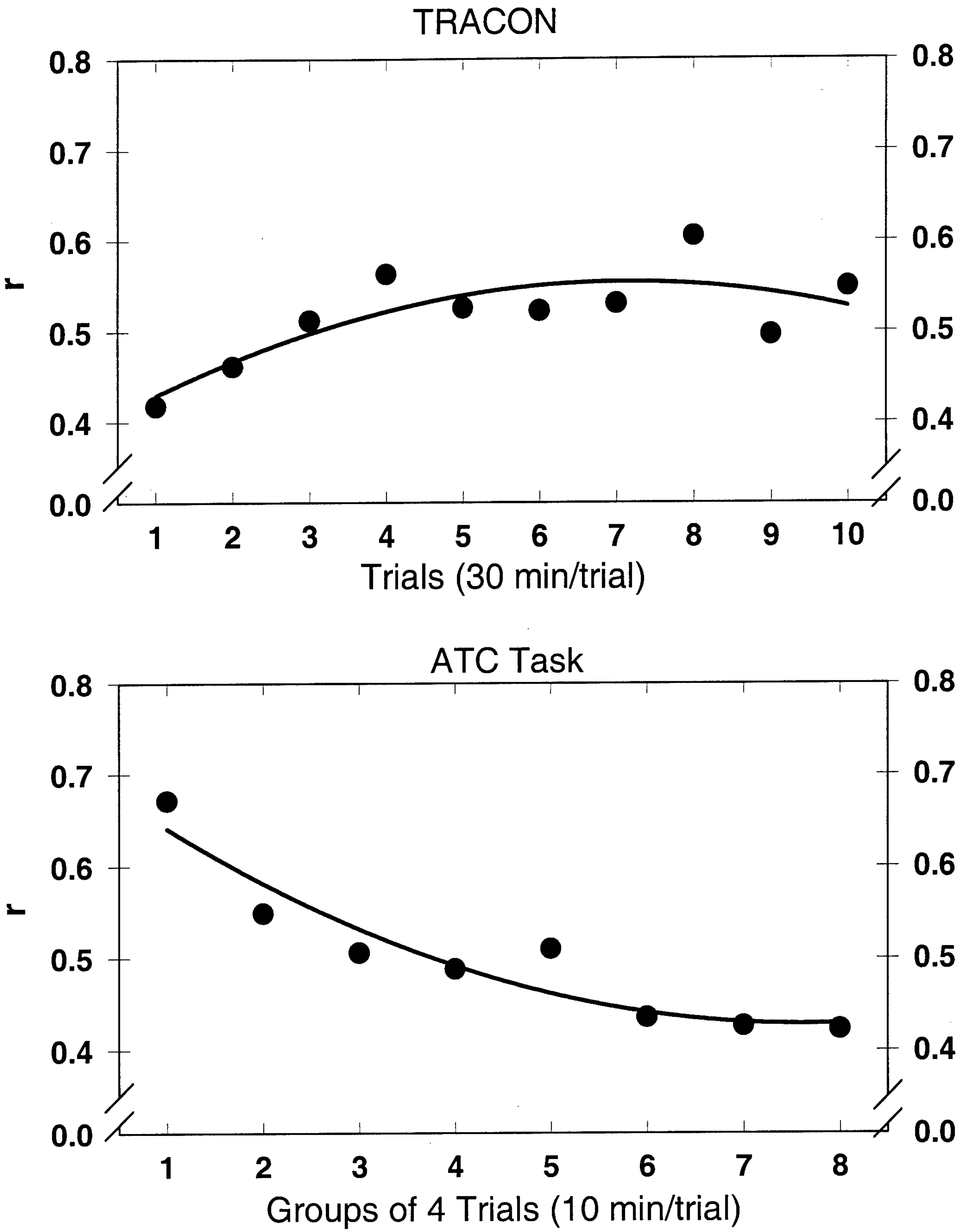


Figure 11. Correlations between general/cognitive ability composite and performance on TRACON(r) and the Kanfer-Ackerman ATC(c) Task.

Predictor measures: Factors and Cross-Correlations. In order to reduce the number of predictor variables, which are composed of 10 cognitive ability measures, 11 perceptual speed variables, and 12 summary psychomotor variables, the first step was to create a set of composite variables across the a priori selected factors of abilities. Thus, composite measures (using unit-weighted *z*-scores) were created for Numerical, Verbal, and Spatial Ability; for the four Perceptual Speed ability factors (Complex, Pattern, Scanning, and Memory), and for the seven identifiable touch-panel psychomotor tasks (Single Tapping, Alternate Tapping, Choice RT, Serial RT, Maze Tracing, Mirror Tracing, and Pursuit), yielding 14 separate predictor composites. To evaluate the underlying structure of these composites, they were subjected to a principal factor analysis. Based on an examination of eigenvalues and a Humphreys-Montanelli Parallel Analysis, four factors were derived. The factors were then subjected to an orthogonal Varimax rotation, which is shown in **Table 11**.

The solution is interesting in that it both supports a priori expectations, and provides an additional validation and illumination of the discussion from the prior experiments and from previous research. The first factor was identified as a broad cognitive ability factor, including Numerical, Verbal, and Spatial ability composites, as well as the Perceptual Speed-Complex composite. (Additional significant loadings on this factor are provided by Perceptual Speed-Memory, Mirror Tracing and Pursuit composites.) The second factor is defined by all of the touch-panel Psychomotor composites, with the exception of the tapping tasks (though the Alternate Tapping task had a salient loading on this factor). The fourth factor is defined by the two Tapping tasks -- Single Tapping and Alternate Tapping. The third factor is defined by the Perceptual Speed abilities, with additional salient loadings by the Perceptual Speed-Complex composite, and the Serial Reaction Time task composite.

Cross-Correlations between Predictors and Criterion Tasks. **Table 12** provides a set of comparisons between the various predictor composites (Cognitive, Perceptual Speed, and Psychomotor Abilities) and the initial and final (post-practice) performance measures on TRACON and the ATC task. In addition, the table lists the cross-correlations between the two criterion tasks, and between initial and final performance for each task. Several noteworthy findings are revealed in this table. First, initial and final performance for TRACON are correlated $r = .500$, which indicates that 25% of the final task performance variance is accounted for by individual differences on TRACON initial performance. The analogous comparison for the ATC task is $r = .643$, indicating that 41% of final task performance is accounted for by individual differences on ATC initial performance. Also, final task performance on TRACON has more in common with both initial ATC task performance ($r = .565$) and final ATC task performance ($r = .474$) than does Trial 1 performance on TRACON ($r = .394$ and .303 respectively).

The raw correlations shown in the first four columns of the Table 12, show, first of all, that cognitive ability measures are significantly and substantially related to both TRACON and ATC task performance. Of the Perceptual Speed abilities, PS-Complex and PS-Memory are most highly related to performance on both criterion tasks, while PS-Pattern and PS-Scanning are significantly related to performance on the ATC task only, as expected

Table 11. Experiment 4. Ability Test Composite Factor Solution -- Varimax Rotation, and Scale Intercorrelations

| Composite | Factor[a] I | II | III | IV | Composite 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cognitive Abilities | | | | | | | | | | | | | | | | | | |
| 1. Numerical Ability | **.799** | .250 | .174 | -.036 | .709* | | | | | | | | | | | | | |
| 2. Verbal Ability | **.647** | -.026 | .144 | .173 | .503 | .444* | | | | | | | | | | | | |
| 3. Spatial Ability | **.876** | .198 | .037 | .030 | .784 | .559 | .732* | | | | | | | | | | | |
| 4. Perceptual Speed-Complex | **.636** | .256 | **.486** | .125 | .619 | .486 | .623 | .684* | | | | | | | | | | |
| Perceptual Speed Abilities | | | | | | | | | | | | | | | | | | |
| 5. Perceptual Speed - Pattern | .064 | .273 | **.718** | .017 | .237 | .077 | .145 | .451 | .556* | | | | | | | | | |
| 6. Perceptual Speed - Scanning | .133 | .008 | **.758** | -.083 | .268 | .228 | .087 | .465 | .578 | .562* | | | | | | | | |
| 7. Perceptual Speed - Memory | **.466** | .168 | **.597** | .170 | .455 | .388 | .506 | .689 | .532 | .472 | .617* | | | | | | | |
| Psychomotor (Touch Panel) Abilities | | | | | | | | | | | | | | | | | | |
| 8. Single Tapping | .166 | .213 | -.081 | **.614** | .159 | .214 | .200 | .167 | .025 | -.065 | .165 | .260* | | | | | | |
| 9. Alternate Tapping | -.026 | **.435** | .284 | **.449** | .132 | .100 | .092 | .303 | .342 | .130 | .297 | .345 | .399* | | | | | |
| 10. Choice Reaction Time | .066 | **.643** | .161 | .150 | .228 | .123 | .202 | .307 | .218 | .155 | .227 | .240 | .375 | .516* | | | | |
| 11. Serial Reaction Time | .106 | **.704** | **.507** | .085 | .375 | .163 | .273 | .465 | .568 | .348 | .515 | .146 | .525 | .631 | .697* | | | |
| 12. Maze Tracing | .102 | **.609** | .201 | .098 | .304 | .074 | .215 | .311 | .299 | .234 | .250 | .219 | .371 | .470 | .519 | .406* | | |
| 13. Mirror Tracing | **.414** | **.523** | -.085 | .134 | .418 | .246 | .439 | .417 | .174 | -.033 | .246 | .273 | .223 | .296 | .329 | .402 | .496* | |
| 14. Pursuit (Maze & Rotary) | **.437** | **.696** | -.074 | .246 | .476 | .269 | .502 | .479 | .193 | -.023 | .349 | .378 | .359 | .502 | .491 | .447 | .659 | .640* |

-------------------

* Diagonal entries in correlation matrix are squared multiple correlations.

Note: Salient factor loadings over .300 are shown in bold (largest loadings) or italics (other salient loadings), correlations larger than .20 are significant at $p = .05$, correlations greater than .26 are significant at $p = .01$.

[a]Factors were identified as follows: I - Cognitive Ability II - Psychomotor, III - Perceptual Speed, and IV - Psychomotor-Tapping

Table 12. Cross-Correlations between Ability Composites and both initial and final task performance for TRACON.

| | TRACON | | ATC Task | | *t*-test (differences) | |
|---|---|---|---|---|---|---|
| | Trial 1 | Trial 10 | Session1 | Session8 | TRACON | ATC |
| <u>Cognitive Ability</u> | | | | | | |
| 1. Numerical Ability | .337** | .508** | .657** | .408** | -1.95 | **3.81**** |
| 2. Verbal Ability | .313** | .323** | .456** | .305** | -.10 | 1.96 |
| 3. Spatial Ability | .450** | .625** | .660** | .402** | **-2.23*** | **3.96**** |
| 4. Perceptual Speed-Complex | .413** | .503** | .628** | .535** | -1.04 | 1.41 |
| <u>Perceptual Speed Abilities</u> | | | | | | |
| 5. Perceptual Speed- Pattern | .166 | .063 | .242* | .248* | 1.02 | -.07 |
| 6. Perceptual Speed-Scanning | .125 | .074 | .202* | .237* | .50 | -.42 |
| 7. Perceptual Speed-Memory | .332** | .405** | .502** | .396** | -.79 | 1.42 |
| <u>Psychomotor (Touch Panel) Abilities</u> | | | | | | |
| 8. Single Tapping | .214* | .213* | .231* | .184 | .01 | .56 |
| 9. Alternate Tapping | -.046 | .160 | .191 | .256* | **-2.06*** | -.78 |
| 10. Choice Reaction Time | .120 | .263* | .387** | .458** | -1.44 | -.93 |
| 11. Serial Reaction Time | .220* | .256* | .438** | .501** | -.36 | -.85 |
| 12. Maze Tracing | .193 | .290** | .278** | .290** | -.99 | -.14 |
| 13. Mirror Tracing | .364** | .414** | .494** | .428** | -.55 | .89 |
| 14. Pursuit (Maze & Rotary) | .345** | .442** | .546** | .487** | -1.07 | .83 |
| <u>Criterion Measures</u> | | | | | | |
| 1. TRACON - Trial 1 | | .500 | .394 | .303 | | 1.14 |
| 2. TRACON - Trial 10 | .500** | | .565 | .474 | | 1.29 |
| 3. K-A ATC Session 1 | .394** | .565** | | | **-2.04*** | |
| 4. K-A ATC Session 8 | .303** | .474** | .643** | | -1.90 | |

NOTE: *N* (98), *t*-test *df*=95. For *t*-test, negative *t* value indicates <u>increase</u> in correlation from Trial 1 /Session 1 to Trial 10/Session 8, positive *t* value indicates <u>decrease</u> in correlation from from Trial 1 /Session 1 to Trial 10/Session 8; * $p < .05$., ** $p < .01$

from the consistent nature of the PS-test constraints, illuminated in the taxonomic work on PS abilities (Ackerman & Rolfhus, 1996). Psychomotor abilities were generally more highly correlated with the ATC task performance, again consistent with expectations. Higher correlations were found for the more complex psychomotor measures (Mirror Tracing and Pursuit) than for the simpler measures (e.g., Tapping). Choice RT and Serial RT ability measures were significantly and substantially related to ATC performance at the end of practice.

In order to evaluate the nature of changes in the predictive validity of the various measures for TRACON and the ATC task over practice, *t*-tests comparing the correlations with initial and final task scores were computed, and are shown in the final two columns of the table. The results of these comparisons for the Cognitive Ability measures illuminate the findings for the general/cognitive ability composite shown in Figure 11, with TRACON showing increasing correlations with general/cognitive ability over practice, and the ATC task showing decreasing correlations with general/cognitive ability over practice. Perceptual Speed abilities showed a generally stable set of correlations across task practice on both TRACON and the ATC task, while the Psychomotor Abilities showed increasing correlations for TRACON (from a very low level to a moderate level of correlations), and showed generally stable correlations (at relatively higher levels) with the ATC task.

Multiple Correlations among Predictor and Criterion Tasks. While the individual correlations shown in Table 12 provide important ability predictor—task performance comparisons, bivariate correlations do not allow for an assessment of the presence of overlapping variance (communality) among the various abilities in predicting individual differences in criterion performance. Multivariate procedures, such as multiple correlation and regression provide one method for evaluating the relative influence specific predictor variables have, in the context of other predictor variables. In order to assess the utility of the new touch-panel based measures of psychomotor abilities (in the context of extant measures of cognitive and perceptual speed ability) a series of multiple correlations were computed.

**Table 13** shows a simplified set of three-step multiple correlations for initial and final criterion task performance measures. In the first step, the Cognitive ability composite measures were used to predict criterion task performance. In the second step, Perceptual Speed composite measures were added to the equation, thus yielding an assessment of the "incremental predictive validity" of these measures. In the third and final step, the new touch-panel Psychomotor ability measures (after removal of the tapping measures, which showed the smallest raw correlations with criterion task performance) were added to the equation -- yielding an assessment of the incremental predictive validity of these abilities. In this manner, the final multiple correlation step indicates whether the new Psychomotor tests add significant independent predictive validity, after the extensive battery of extant paper-and-pencil based measures are allowed to account for all of the common variance among the various predictor measures. This is essentially the most conservative test for assessing the utility of the psychomotor measures, and is shown as "Method 1."

Table 13. Multiple correlations for predicting Knowledge Scale composite scores

| Method 1 | | Step 1 Cognitive Ability[a] | Step 2 Perceptual Speed[b] | Step 3 Psychomotor[c] |
|---|---|---|---|---|
| TRACON Trial 1 | $R^2$ to add | .210** | .032ns | .036ns |
| | Total $R^2$ | .210** | .242** | .278** |
| TRACON Trial 10 | $R^2$ to add | .393** | .044ns | .034ns |
| | Total $R^2$ | .393** | .437** | .471** |
| K-A ATC Session 1 | $R^2$ to add | .488** | .056* | .059* |
| | Total $R^2$ | .488** | .544** | .603** |
| K-A ATC Session 8 | $R^2$ to add | .187** | .111* | .147** |
| | Total $R^2$ | .187** | .298** | .445** |

| Method 2 | | Step 1 Psychomotor[c] | Step 2 Perceptual Speed[b] | Step 3 Cognitive Ability[a] |
|---|---|---|---|---|
| TRACON Trial 1 | $R^2$ to add | .164** | .067ns | .046ns |
| | Total $R^2$ | .164** | .231** | .278** |
| TRACON Trial 10 | $R^2$ to add | .229** | .158** | .084** |
| | Total $R^2$ | .229** | .387** | .471** |
| K-A ATC Session 1 | $R^2$ to add | .376** | .144** | .083** |
| | Total $R^2$ | .376** | .520** | .603** |
| K-A ATC Session 8 | $R^2$ to add | .368** | .076* | .001ns |
| | Total $R^2$ | .368** | .444** | .445** |
| Degrees of freedom | | | | |
| | Numerator | 3 | (4) 7 | (5) 12 |
| | Denominator | 90 | 86 | 81 |

-----------------

Notes: ns = not significant; *$p$ < .05; ** $p$ < .01.

[a] Cognitive Abilities included (Numerical, Verbal, and Spatial composites)

[b] Perceptual Speed Abilities included (PS-Complex, PS-Pattern, PS-Scanning, and PS-Memory)

[c] Psychomotor Abilities included (Choice RT, Serial RT, Maze Tracing, Mirror Tracing, and Pursuit)

Although the total sample of $N = 98$ is modest for the kind of analysis reported here, the results are concordant with a priori expectations, and quite impressive for the approach adopted in this project. Psychomotor measures were expected to be minimally "independently" associated with performance on the complex problem solving and decision-making processes involved in the TRACON task (i.e., once common variance associated with Cognitive and Perceptual Speed abilities was accounted for). Indeed, for both initial and final TRACON performance, the Psychomotor measures only accounted for 3-4% of the total variance, after Cognitive ability and Perceptual Speed were entered into the regression equation. Cognitive ability increased in predictive validity from 21% of variance accounted for in Trial 1, to 39% of the variance in Trial 10, while Perceptual Speed remained relatively constant in predictive utility, around 3-4% of the variance. In all, 28% and 47% of the variance in TRACON task performance was accounted for by all types of ability measures, at initial and final task performance, respectively.

In contrast, Cognitive ability accounted for 49% of variance in initial ATC performance, but only 19% of the variance in final ATC task performance. Concordant with previous research, Perceptual Speed increased in predictive validity from accounting for 6% of the initial performance variance to about 11% of final task performance variance. Most notably, Psychomotor abilities were expected to make their most significant contribution in predicting final ATC task practice. Indeed, Psychomotor abilities increased in variance accounted for, from initial ATC performance (6% of variance) to final ATC performance (15% of variance). In all, although total variance accounted for by the various predictors declined from 60% to 44%, it is critical to recall that there was simply less individual-differences variance available to be accounted for, in that between-subject variability declined 17% over task practice. The increasing amount of variance accounted for by the Psychomotor ability measures, even in the context of decreasing variability in task performance, is a substantial and impressive result for these new tests. Moreover, it is important to keep in mind that the total amount of time accorded to psychomotor testing was substantially less than the amount of time required to administer the paper-and-pencil cognitive ability and psychomotor tests. This issue will be discussed in some detail in the conclusions section below.

Although the questions related to the implementation of the new touch-panel measures of Psychomotor abilities are generally answered by the Method 1 multiple correlation analysis shown in the table, it is instructive to examine the common variance among the predictors by an alternative route -- that is by performing essentially a "bottom-up" approach -- where the Psychomotor measures are entered into the prediction equation first, followed by the Perceptual Speed measures, and finally by the Cognitive ability measures. The results of these analysis are shown as "Method 2." Interestingly, the psychomotor measures, when entered first into the prediction equations, are significant and substantial predictors of criterion task performance in their own right. In all but TRACON Trial 1, Perceptual Speed measures provide incremental predictive validity over the Psychomotor measures. Most importantly though, the Cognitive ability measures provide essentially zero (.001) incremental predictive validity over the Psychomotor and Perceptual Speed measures on the

final session of the ATC task -- again supporting the assertion of the important predictive properties of the Psychomotor and Perceptual Speed abilities in predicting performance on tasks that allow for the development of automatized information processing to accomplish task performance.

Experiment 4 Discussion

Overall, the results of Experiment 4 further demonstrate the unique associations of the new Psychomotor Ability measures, in the context of Cognitive and Perceptual Speed ability measures. In addition, the results are highly supportive of the theoretical predictions made for Perceptual Speed and Psychomotor abilities in accounting for criterion task performance. Finally, the touch-panel instantiation of the Psychomotor ability measures were demonstrated to have significant incremental predictive validity for final (skilled) performance on the ATC criterion task -- even when overall between-subject variance declined with task practice. While general ability correlates of task performance decreased in the ATC task, the Psychomotor (and Perceptual Speed) measures showed an increase in predictive validity -- confirming the claim made by Adams (1957) and later refined by Ackerman (1988, 1989, 1990) that prediction of skilled task performance is indeed possible based on ability measures that are administered *prior* to task practice. These results *refute* the assertions by others (e.g., Fleishman, 1972; Hulin, Henry, & Noon, 1990) that declining predictive validity is necessary over task practice, and point to important methods for improving selection for skilled performers.

## VIII. Conclusions/Future Directions

During the past 50 years of psychological testing research and practice, psychomotor testing has been relegated a category of 'valid, but impractical' predictors of many kinds of real-world behaviors, particularly in the domain of work performance. While such measures are still used to a substantial degree in neurological assessment, where the testing is typically performed in one-to-one examiner/examinee format, both group testing and testing at remote sites have been avoided because of the logistical difficulties outlined by Melton in 1947. The current project results suggest that the adaptation of some key psychomotor tests to the computerized touch-panel system can effectively eliminate all of these barriers to testing. Indeed, with touch-panel computer monitors now ubiquitous in the consumer markets (e.g., automated teller machines and kiosks) and the service industry (point of sale computers, restaurant entry systems), obtaining and using such devices has become quite straightforward. The results indicate that even the more sophisticated TouchPen design (over the more popular finger entry system), while preferred by the examinees, makes relatively little difference in the validity of test results.

From a research perspective, the computerized touchpanel removes additional barriers toward investigating psychomotor abilities. In the past, researchers have had to endure trips down to the 'tool shop' to have any changes (e.g., target size) made to existing psychomotor apparatus tests. With a computerized system such as the one described here, a simple scripting program can allow for changes to such parameters to be made in a matter of

minutes. With such a system, it is possible for paradigmatic research on the characteristics of psychomotor tests to be investigated with little additional investment.

It is important to note that one major advantage of the new touch-panel psychomotor tests is their brevity -- each test typically takes less than 5-10 minutes of administration time, in comparison to cognitive tests (which often take as long as 20-40 minutes to administer). While the perceptual speed tests that have been developed under AFOSR sponsorship are similar in administration time to the psychomotor tests (typically about 6-8 minutes testing time), they have the shortcoming of being cumbersome to score -- involving tedious human scoring procedures. When the validity of the psychomotor tests is comparable to the cognitive and/or perceptual speed tests, the touch-panel instantiation is clearly the most desirable in terms of time and effort in testing and scoring.

One major challenge remains for research and validation exist at this point in the research program, that of validation to more job-relevant performance. A research study that partly addresses this issue was completed during the grant period (Ackerman, Cianciolo, & Bowen, in progress). In this study, several of the perceptual speed and touch-panel psychomotor tests used in the current project were administered to School of Dentistry students. Although the data are currently being subjected to detailed analysis, preliminary results suggest that the Maze/Mirror Tracing and Maze/Rotary Pursuit psychomotor tests provided significant incremental validity (beyond extant selection measures of cognitive and perceptual abilities) in predicting course grades for these students. However, more in-depth validation studies will be needed to more fully evaluate the ultimate utility of these promising ability assessment procedures.

All in all, this project was highly successful and all of the research program goals were satisfied. The seven touch-panel psychomotor tests were developed and validated within the laboratory confines allowed in the research program. The tests are easily portable to standard personal computer systems, and are fully ready to be adapted for real-world validation purposes. It is hoped that future investigations will extend the validation work conducted to date, and that these tests might also be used in other situations that call for assessment of psychomotor abilities, such as basic research in aging or in neurological assessment.

## IX. References


Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10*, 101-139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*, 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*, 288-318.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 883-901.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77*, 598-614.

Ackerman, P. L. (1996, August). *Psychomotor/perceptual speed abilities: New measurement techniques and trait overlap.* Paper presented at the 1996 annual meeting of the American Psychological Association. Toronto, Canada.

Ackerman, P. L., Cianciolo, A. T., & Bowen K. R. (in progress). Touch-panel monitor based psychomotor tests for predicting skilled performance: An exploratory study with School of Dentistry students. Paper to be submitted for presentation at the 1999 Human Factors and Ergonomics Society annual meeting.

Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78,* 413-432.

Ackerman, P. L. & Kanfer, R. (1994). *Kanfer-Ackerman Air Traffic Controller Task© CD-ROM Database, Data Collection Program, and Playback Program* and *Manual.* Minneapolis: University of Minnesota.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1*, 270-304.

Ackerman, P. L., & Kyllonen, P. C. (1991). Trainee characteristics. Chapter in J. E. Morrison (Ed.) *Training for performance: Principles of applied human learning* (pp. 193-229). West Sussex, England: John Wiley & Sons, Ltd.

Ackerman, P. L., & Lohman, D. F. (1990). *An investigation of the effect of practice on the validity of spatial tests.* Final report to the Navy Personnel Research & Development Center. (Contract #N66001-88-C-0291). San Diego, CA: Author.

Ackerman, P. L., & Rolfhus, E. L. (1996, November). *A Taxonomic Study of Perceptual Speed Abilities.* Paper presented at the 1996 annual meeting of the Psychonomic Society. Chicago, IL.

Ackerman, P. L., & Woltz, D.J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, and volition. *Journal of Educational Psychology, 86*, 487-515.

Adams, J. A. (1957). The relationship between certain measures of ability and the acquisition of a psychomotor criterion response. *The Journal of General Psychology, 56*, 121-134.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*, 41-74.

Adkins, D. C., Primoff, E. S., McAdoo, H. L., Bridges, C. F., & Forer, B. (1947). *Construction and analysis of achievement tests: The development of written and performance tests of achievement for prediction job performance of public personnel.* Washington, DC: U.S. Government Printing Office.

Alexander, J. R., Ammerman, H. L., Fairhurst, W. S., Hostetler, C. M., & Jones, G. W. (1990). FAA air traffic control operations concepts volume VI: ARTCC/Host en route controllers. Washington, D.C.: FAA report no. DOT/FAA/AP-87-01.

Allison, R. B. (1960). *Learning parameters and human abilities.* (Office of Naval Research Technical Report). Princeton, NJ: Educational Testing Service.

Ammerman, H. L., Becker, E. S., Jones, G. W., Tobey, W. K., & Phillips, M. K. (1987). *FAA air traffic control operations concepts volume I: ATC background and analysis methodology*. Washington D.C.: FAA report no. DOT/FAA/AP-87-01.

Anastasi, A. (1982). *Psychological Testing*. New York: Macmillan.

*Army Air Force Aviation Psychology Program Research Reports, Vol 1-19* (1947). Government Printing Office.

Brown, C. W., & Ghiselli, E. E. (1952). The relationship between the predictive power of aptitude tests for trainability and for job proficiency. *Journal of Applied Psychology, 36*, 370-372.

Buckley, E. P., Debaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation*. DOT/FAA/CT-83/26. FAA Technical Note, FAA: Atlantic City Airport, NJ.

Carretta, T. R. (1987). *Basic Attributes Test (BAT) system: The development of an automated test battery for pilot selection.* (AFHRL-TR-87-9, AD-A186 649). Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Cianciolo, A. T. (1997). *Computerized assessment of psychomotor ability*. Unpublished M.A. Thesis. Minneapolis: University of Minnesota.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement, 52*, 721-725.

Field, K. A. (1998). *Assessment and applications of psychomotor abilities using a new computerized (touch-panel) method.* Unpublished Ph.D. Thesis. Minneapolis: University of Minnesota.

Fleishman, E. A. (1953). Testing for psychomotor abilities by means of apparatus tests. *Psychological Bulletin, 50*, 241-263.

Fleishman, E. A. (1954). Dimensional analysis of psychomotor abilities. *Journal of Experimental Psychology*, *48*, 437-454.

Fleishman, E. A. (1956). Psychomotor selection tests: Research and application in the U.S. Air Force. *Personnel Psychology, 9*, 449-467.

Fleishman, E. A. (1958). Dimensional analysis of movement reactions. *Journal of Experimental Psychology*, *55*, 438-453.

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist, 27* (11), 1017-1032.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests.* New York: Wiley.

Goska, R. E., & Ackerman, P. L. (1996). An aptitude-treatment interaction approach to transfer within training. *Journal of Educational Psychology, 88*, 249-259.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Army Air Forces Aviation Psychology Program Research Reports: Printed Classification Tests.* Report No. 5. Washington, DC: U. S. Government Printing Office.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258-348). The Free Press.

Hull, C. L. (1928). *Aptitude Testing.* New York: World Book Company.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin, 107*, 1-13.

Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10*, 193-206.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology - Monograph, 74*, 657-690.

Kanfer, R. & Ackerman, P. L. (1990). *Ability and metacognitive determinants of skill acquisition and transfer.* Final Report to Air Force office of Scientific Research. Minneapolis: University of Minnesota.

Kanfer, R., & Ackerman, P. L. (1996). A self-regulatory skills perspective to reducing cognitive interference. In I. G. Sarason, B. R. Sarason, & G. R. Pierce (Eds.), *Cognitive interference: Theories, methods, and findings* (pp. 153-171). Mahwah, NJ: Erlbaum.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Murray Hill, NJ: Bell Laboratories.

Kyllonen, P. C., and Christal, R. E. (1989). Cognitive modeling of learning abilities: A status report of LAMP. In R. Dillon and J. W. Pellegrino (Eds.), *Testing: Theoretical and applied issues*, San Francisco: Freeman.

Kyllonen, P. C., Tirre, W. C., & Christal, R. E. (1991). Knowledge and processing speed as determinants of associative learning. *Journal of Experimental Psychology: General, 120*, 57-79.

Landon, T. E., & Ackerman, P. L. (1991) *Job Performance for the En Route ATCS: A Review with Applications for ATCS Selection.* Center for Transportation Studies, University of Minnesota Technical Report.

Levine, E. L., Spector, P. E., Menon, S., Narayanan, L., & Cannon-Bowers, J. (1996). Validity generalization for cognitive, psychomotor, and perceptual tests for craft jobs in the utility industry. *Human performance, 9*, 1-22.

McDermid, C. & Smith, K. U. (1964). Compensatory reaction to angularly displaced visual feedback in behavior. *Journal of Applied Psychology, 48*, 63-68.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107-127.
Means, B., Mumaw, R., Roth, C., Schlager, M., McWilliams, E., Gagné, E., Rice, V., Rosenthal, D. & Heon, S. (1988). *ATC training analysis study: design of the next-generation ATC training system.* Technical report OPM/342-036. Washington, DC: FAA, Office of Training and Higher Education.
Melton, A. W. (Ed.). (1947). *Army Air Forces Aviation Psychology Program Research Reports: Apparatus Tests.* Report No. 4. Washington, DC: U.S. Government Printing Office.
Montanelli, R. G., Jr., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika, 41,* 341-348.
Münsterberg, H. (1913). *Psychology and industrial efficiency.* Boston: Houghton Mifflin.
Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.
Ree, M. J., Earles, J. A., & Teachout, M. S. (1992). *General cognitive ability predicts job performance.* (AL-TP-1991-0057, AD-A245 099), ERIC #ED361600. Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower & Personnel Division.
Reynolds, B. (1952a). The effect of learning on the predictability of psychomotor performance. *Journal of Experimental Psychology, 44,* 189-198.
Reynolds, B. (1952b). Correlations between two psychomotor tasks as a function of distribution of practice on the first. *Journal of Experimental Psychology, 43,* 341-348.
Salvendy, G., & Seymour, W. D. (1973). *Prediction and development of industrial work performance.* New York: Wiley.
Seashore, R. H. (1940). An experimental and theoretical analysis of fine motor skills. *American Journal of Psychology, 53,* 86-98.
Seashore, R. H., Buxton, C. E., & McCollom, I. N. (1940). Multiple factorial analysis of fine motor skills. *American Journal of Psychology, 53,* 251-259.
Snoddy, G. S. (1920). An experimental analysis of a case of trial and error learning in the human participant. *Psychological Monographs, 28* (2, Whole No. 124).
Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47-103). Hillsdale, NJ: Erlbaum.
Tomporowski, P. D., Simpson, R. G., & Hager, L. (1993). Method of recruiting subjects and performance on cognitive tests. *American Journal of Psychology, 106,* 499-521.
Tucker, L. R, & Finkbeiner, C. T.(1981). *Transformation of factors by artificial personal probability functions.* (Report #RR-81-58). Princeton, NJ: Educational Testing Service.
Wolfe, J. H. (1997). Incremental validity of ECAT battery factors. *Military Psychology, 9,* 49-76.
Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117,* 319-331.

## Appendix A. The Kanfer-Ackerman Air Traffic Control (ATC) Task

The ATC task is a rule-based, real-time, computer-driven task that simulates some of the activities performed by air-traffic controllers. The overall objective for subjects who perform this task is to land planes safely and efficiently. An example of the ATC task display is presented in the figure below. As shown, the following task elements are displayed when performing the task: (a) four runways, (b) 12 hold pattern positions, and (c) a queue stack with asterisks indicating planes requesting permission to enter the hold pattern. Two runways run North-South; two runways run East-West. One North-South and one East-West runway is short; one North-South and one East-West runway is long.

```
 FLT#     TYPE     FUEL     POS.
 ----     ----     ----     ----       Score    :  150
                             3 n       Landing Pts:  150   Penalty Pts:     0
                             3 s       Runways : DRY
  161      747        5      3 e       Wind     : 40 - 50 knots from SOUTH
                             3 w
  403      747        6      2 n      +--------------------------------------+
  889      727        6      2 s      | Flts in Queue: ***                   |
                             2 e      |        <F1> to accept                |
                             2 w      +--------------------------------------+
  631      727        6      1 n      +--------------------------------------+
  144      prop       5      1 s      |          Winds 40-50 knots           |
  903      DC10       6      1 e      |           Winds from South           |
  122      747     *  3      1 w      |             Runways dry              |
                                      +--------------------------------------+
========================= s    #1 <-

=========727======        s    #2

||||||||||||||||||||||||| e    #3  +--------------------------------------+
                                   |      Can use short runways when:     |
|||||||||||||||||         e    #4  |    747 - Never         Prop - Always |
                                   |   DC10 - Not Icy & not 40-50 knots   |
                                   |        727 - Dry or 0-20 knots       |
                                   +--------------------------------------+
```

Figure A1. Sample screen from the Kanfer-Ackerman Air Traffic Control© Task.

The hold pattern, located in the middle right section of Figure A1, contains twelve hold pattern positions, divided into three levels (analogous to three platters at different altitudes in the sky over the airport). Hold pattern position is indicated by number and letter in the Position (POS) column. Level 1 hold positions represent the lowest altitude and Level 3 hold positions represent the highest altitude. Four positions, corresponding to the points of the compass (i.e., N, S, E, W), are available in each level.

Planes are admitted to the hold pattern from the queue. The queue, located at the upper right of the screen, displays number of planes requesting permission to enter the hold pattern. Each plane request is represented by an asterisk. Planes enter the queue at the rate of one approximately every 7 seconds. Plane requests remain in the queue until the subject places the plane in the hold pattern.

As shown in the figure, four types of planes enter the subject's hold pattern; 747's, 727's, DC10's, and Props. Plane information is displayed in the hold pattern. When a plane is placed in the hold pattern, Flight Number (FLT#), Plane Type (TYPE), and Number of Minutes of Fuel remaining (FUEL) are displayed. Within each trial an approximately equal number of plane types is randomly drawn from the queue. Fuel remaining, determined when the plane was brought into the hold pattern, are randomly varied from four to six minutes. Once the planes enter the hold pattern, fuel remaining decreased in real time, such that when zero minutes of fuel remained, the plane crashes.

Subjects also receive information on airport weather conditions. Weather information is used (in accordance with the rule set) to determine what planes were allowed to land on which runways. Three elements comprised weather conditions; wind speed, wind direction, and ground condition. Wind speed and wind direction information are displayed on the "wind" line at the top right corner of the screen. Ground condition is displayed on the "runways" line. Updates to weather conditions are displayed throughout each task trial. Three types of wind speed are presented (0 - 20 knots, 25 - 35 knots, and 40 - 50 knots). Four types of wind direction are displayed (North, South, East, and West). Three levels of ground conditions are used (runways dry, wet, or icy). Changes in weather conditions (defined as a change in at least one of the three weather condition components) are varied randomly during a task trial. On average, these changes occur about twice a minute (i.e., 20 weather changes are initiated during each 10-minute task trial).

Feedback and Knowledge of Results. The first component of knowledge of results was a one-to-one mapping between keystrokes made by the subject, and operation of a cursor on the screen. As planes are selected, various parts of the display are highlighted. When a plane is moved from one hold position to another, or to a runway, the subject sees an analogous change to the display. Subjects also receive three types of continuously updated performance information throughout each trial. Cumulative performance (Score) for the current trial is based upon a specified point scheme. Subjects receive 50 points for each plane successfully landed. Ten points are deducted for each technical error made (violation of the rules). One hundred points are deducted from the performance score for each plane that runs out of fuel in the hold pattern (i.e., plane crashes). Performance scores may be negative or positive depending on how many planes are landed, relative to number of errors made and planes crashed. In addition, subjects receive separate landing (Landing Pts.) and error (Penalty Pts.) information. Landing Pts., based upon the number of planes landed, started at zero and increase by 50 points for each plane landed. Penalty Pts., reflecting the number of rule violations and plane crashes, start at zero and decrease for each error.

Task requirements. Subjects perform three principal actions: (a) accepting planes into the hold pattern, (b) moving planes in the three-level hold pattern, and (c) landing planes on appropriate runways. Subjects manipulate planes using only four keys on the computer keyboard (plus keys for rule call-ups). For example, planes are moved down the hold pattern by pressing the "down-arrow" (↓) key once for each position in the hold pattern. A one-to-one correspondence between keyboard and screen actions is maintained by linking

each keyboard response to movement of a small cursor arrow on the screen. Specific keyboard actions taken to move a plane in the hold pattern and to place a plane on a runway result in highlighting of the target plane and real-time movement of the plane across the runway. Successful performance on this task requires knowledge about how to make plane movements using the computer keyboard as well as knowledge of the rules governing plane movements and landings.

## Appendix B. Terminal Radar Approach Control (TRACON) Simulation Task

The task used for this research is a uniquely modified Professional version (V1.52) of TRACON simulation software, developed by Wesson International. Versions of this program have been, and are, in use in several locations in this country (including the FAA, NASA, DoD, and several colleges and university airway sciences programs) for training of air traffic controllers. Modifications for the current instantiation of the program allowed for the collection of a variety of data, described in more detail below.

Task analysis is a critical component of empirical investigation with any task, but especially with a complex task such as TRACON. In fact, one of the salient reasons for adoption of the TRACON task here is that there has been extensive historical work on task analysis for the full range of Air Traffic Controller tasks, including the subset implemented in TRACON. In particular, a five-volume set of task analysis materials has been prepared by the FAA (e.g., Alexander, et al., 1990; Ammerman, et al. 1987). Additional task analyses have been conducted by the U.S. Air Force, and more recently by HumRRO (Means et al., 1988). All of these have been carefully studied in an effort to derive the critical components of the simulation task adopted for empirical evaluation (see, e.g., Landon & Ackerman, 1991).

The task requires that subjects learn a set of rules for positive air traffic control, including (a) reading flight strips, (b) declarative knowledge about radar beacons, airport locations, airport tower handoff procedures, en-route center handoff procedures, (c) plane separation rules and procedures, (d) monitoring strategies, and (e) strategies for sequencing planes for maximum efficient and safe sector traversal. In addition, subjects are required to acquire human-computer interface skills: including issuing mouse-based commands, menu retrieval, keyboard operations, and integration between visual and auditory information channels. Although the task represents a substantial reduction of rules and operational demands in comparison to the real-world job of an Air Traffic Controller (ATC), it represents an excellent simulation vehicle for study of skill acquisition, within a time-frame that can be handled in a laboratory-based research environment.

Display. TRACON presents the controller (subject) with a simulated color radar screen, depicting a region of airspace, radio navigational tower locations (VOR), airports, sector boundaries, and range rings. Planes are identified by an icon on the radar scope, with a data tag that indicates plane identification and altitude information. In addition, two sets of "Flight Strips" are presented at the right side of the display, a "Pending" and an "Active" set. Each flight strip contains information about a particular flight, including identification information, plane type, requested speed and altitude, and sector entry and exit destination information (See Figure B1). Finally, at the bottom of the screen is a "Communications Box" -- which shows commands issued to planes (and responses by the pilots), along with the controller's "score" for the current simulation.

When planes are about to enter the subject's sector (at a boundary or on the runway of an airport) this information is announced over the headset (e.g., "Northwest 123 ready for takeoff," or "Delta 123, with you, level at 9,000"). However, no flight is allowed to cross the sector boundary or take off from an airport without explicit authorization.

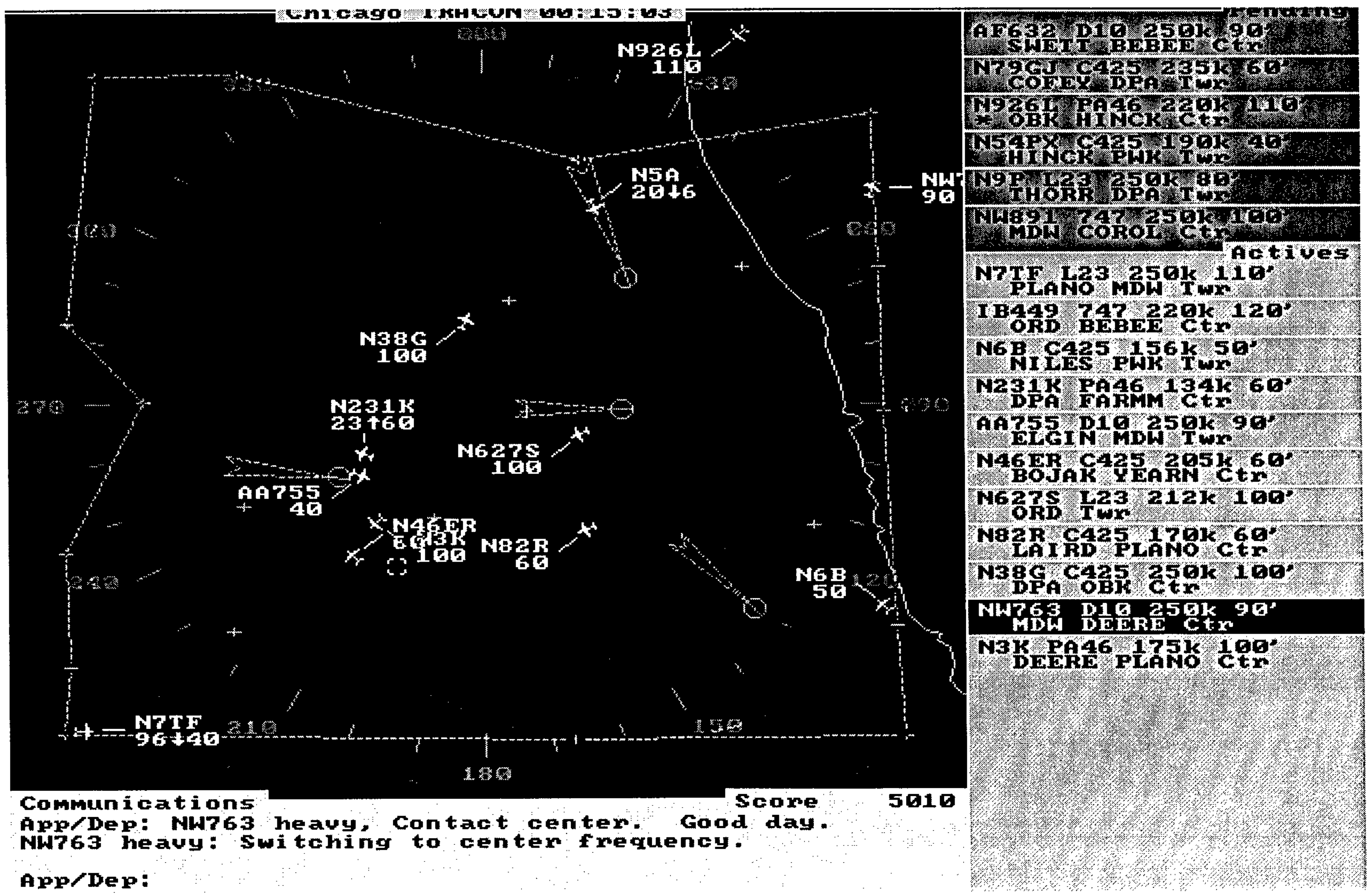
Chicago TRACON 00:15:03
Pending
AF632 D10 250k 90'
SWEIT BEBEE Ctr
N79GJ C425 235k 60'
COFEY DPA Twr
N926L PA46 220k 110'
OBK HINCK Ctr
N54PX C425 190k 40'
HINCK PWK Twr
N9P L23 250k 80'
THORR DPA Twr
NW891 747 250k 100'
MDW COROL Ctr
Actives
N7TF L23 250k 110'
PLANO MDW Twr
IB449 747 220k 120'
ORD BEBEE Ctr
N6B C425 156k 50'
NILES PWK Twr
N231K PA46 134k 60'
DPA FARMM Ctr
AA755 D10 250k 90'
ELGIN MDW Twr
N46ER C425 205k 60'
BOJAK YEARN Ctr
N627S L23 212k 100'
ORD Twr
N82R C425 170k 60'
LAIRD PLANO Ctr
N38G C425 250k 100'
DPA OBK Ctr
NW763 D10 250k 90'
MDW DEERE Ctr
N3K PA46 175k 100'
DEERE PLANO Ctr
N926L 110
N5A 20↓6
N38G 100
N231K 23↑60
N627S 100
AA755 40
N46ER 60
N3K 100
N82R 60
N6B 50
N7TF 96↓40
NW 90
270
180
210
150
Communications
Score 5010
App/Dep: NW763 heavy, Contact center. Good day.
NW763 heavy: Switching to center frequency.
App/Dep:

Figure B1 (Preceding Page). Static copy of TRACON® screen. There are three major components to the display. The right hand side of the screen shows Pending (not under control) and Active (under control) flight strips. Each flight strip lists (a) plane identifier, (b) plane type, (c) requested speed, (d) requested altitude, (e) Radar fix of sector entry, (f) Radar fix of sector exit (including Tower or Center). The lower part of the screen shows a communications box that gives a printout of the current (and last few) commands issued by the subject, and the responses from pilots or other controllers. The main part of the screen shows a radar representation of the Chicago sector. Planes are represented by a plane icon, and a data tag (which gives the identifier, the altitude, and an indication of current changes in altitude). The sector is bounded by the irregular dotted polygon describing a perimeter. Radar fixes are shown as small (+) figures on the radar screen. Airports are shown with approach cones, and a circle indicating the facility proper. A continuous radar sweep is shown (updating at 12 o'clock, every 5 sec). Range rings are also displayed, indicating 5 mile distances.

Task Controls and Knowledge of Results. Subjects interact with the TRACON simulation in several ways. A mouse is used for the majority of input activities, although the keyboard was also used alone, or in conjunction with the mouse. For each plane command, a menu of command choices is displayed on the screen. For turns (Left or Right), a small wheel was shown so that a direction was selected by pointing to the appropriate place on the wheel, and depressing the mouse key. Altitude and Speed commands resulted in the display of a small linear display from which the subject selected a particular value. Direct and Hold commands require the subject to move the mouse-cursor to a specific VOR fix or airport, and select a location. Resume and Handoff commands have no additional menus, but are initiated directly.

Additional commands for information (Flight Path, Plane Type, and Plane Current Heading (in degrees) and Airspeed (in knots) may also be obtained. Information pertaining to the sector constraints (Map of VOR/Airport fixes; and Airport Information, including final approach heading and altitude requirements) may be called up with keyboard commands.

Knowledge of results is provided visually (by text in the communications box) and auditorially with a read-back by the pilot or other controller (using digitized speech broadcast over the subject's headset). If a command is not allowed (e.g., asking a pilot to increase or decrease speed beyond the limitations imposed by the type of plane), the visual and auditory response indicated a failure to comply with the command (e.g., "Sorry, but that is below my 'stall' speed!"). Handoff commands differ from the other commands, in that a handoff to another sector is only accepted when the plane is within 5 miles of the sector boundary. All other requests for handoff are refused by Center Controllers.

In addition, planes follow (as nearly as possible) the commands issued by the subject. Turn, altitude change, and speed change commands are processed by the computer, and are carried out in accordance with the limitations imposed by each aircraft type (e.g., smaller planes turned in a smaller radius than Boeing 747's, but 747's climbed more quickly than the smaller planes). Each plane performs within the constraints that were displayed when a subject calls up the information for that plane type.

Finally, when errors occur (e.g., separation conflicts, near misses, crashes, missed approaches, handoff errors), additional information is presented to the subject. In each of these cases, an alert circle around the plane(s) in question is presented on the screen, and a series of tones are presented over the headset. If two planes crash, a message appeared on the screen indicating which of the planes crashed. (In normal training, the simulation is immediately halted under such conditions. However, because it is not desirable to minimize learning opportunities of subjects who have crashes, the simulation continues under such circumstances.)

Points. Subjects are told to perform the task so that they maximize successful disposition of all flight paths, but that safety is a critical component of the task. Points are given for successful accomplishment of each plane's flight plan, and penalty points are deducted for both commission or omission errors. Points assigned are based on a priori judgements of task component difficulty (e.g., arrivals were more difficult to accomplish than overflights, so arrivals received three times as many points). The point assignments are

used to encourage subjects to develop an appropriate strategy for task component emphasis.

Trial Description Trials for the task are created and pretested to be roughly equivalent in difficulty. Each trial contains planes that are divided into three basic categories (Overflights, Departures, and Arrivals). Overflights are planes that enter and exit the subject's airspace at cruising altitudes. Subjects are required to acknowledge these airplanes as they approach a boundary VOR fix, monitor progress through the sector, and handoff to a "Center" controller. Departures are planes that originate at one of the four airports, climb to a cruising altitude and are handed off to a "Center" controller. Subjects are required to release departures from airports, evaluate and remediate potential conflicts as the planes climb to a cruising altitude and turn to intercept their intended flight paths, and then handoff planes to the appropriate Center controller. Arrivals enter the subject's airspace from one of the boundary VOR fixes, and have to be landed at a designated airport. Subjects are required to direct arrivals onto an appropriate heading and altitude to provide an acceptable handoff to the appropriate Tower controller, then these planes can land. Practice flights, which originate at an airport, but have to be correctly vectored to be landed again at the same airport are classified as "Arrivals," because demands of these flights are most similar to other arrivals. For all flights, the subject is required to maintain legal separation (at least 1000 ft in altitude, or 3 mi horizontally).

Each trial is comprised of 16 overflights and departures (with roughly equal frequency), and 12 arrivals. The planes request entry to the airspace at irregular intervals that are constrained to require the subject to be always occupied with at least one active target. The trials are also constrained so that perfect performance (handling all 28 planes successfully) is just beyond the skill level achieved by subject matter experts. Each trial is concluded in 43 - 45 min. That is, in order to provide equivalent practice time across subjects, the trials are ended with time constraints, rather than waiting until all planes are handled -- which otherwise introduces a substantial variance in practice time.

A successful "handle" of a flight is the appropriate accomplishment of the respective flight plan. That is, for a departure or an overflight, the accomplishment was a successful handoff to the appropriate Center controller. For a landing, the accomplishment was the successful landing of the airplane.

Errors in performance take a variety of different forms, as follows: (1) Incorrect speed or altitude for center handoff; (2) A failure to handoff the plane; (3) For arrival flights, errors include "wrong approach altitude," or "wrong approach heading" (which requires the subject to reorient the plane for another landing attempt); and (4) Separation conflicts, near misses, and crashes (for a differing degrees of airspace proximity violations).

Performance Measures. After extensive review of the raw data from initial experiments (which include every command issued by subjects during TRACON trials, and a series of summary data for each airplane in a simulation, and for each simulation overall), a general criterion of merit has been selected that reflected overall task performance. This measure, called "Overall Performance" is computed as the sum of all flights accepted into the sector that have a final disposition within the simulation time (minus any planes that are incorrectly disposed of -- e.g., crashes, not-handed-off, vectored off the radar screen). This

measure is generally concordant with results from the examination of the criterion space for FAA ATC simulation research (e.g., see Buckley, Debaryshe, Hitchner, & Kohn, 1983). Other measures are also computed, to reflect declarative knowledge (information requests) and task component processing (separate scores for number of arrivals, departures, and overflights accepted into the sector).